

UNIVERSITY OF ALBERTA

STRUCTURAL STUDIES OF THE INTERACTIONS BETWEEN
SERINE PROTEINASES AND PROTEIN INHIBITORS

by

KUI HUANG



A THESIS

SUBMITTED TO THE FACULTY OF GRADUATE STUDIES AND RESEARCH
IN PARTIAL FULFILLMENT OF THE REQUIREMENTS FOR THE
DEGREE OF DOCTOR OF PHILOSOPHY

DEPARTMENT OF BIOCHEMISTRY

EDMONTON, ALBERTA

FALL 1995

7



National Library
of Canada

Acquisitions and
Bibliographic Services

395 Wellington Street
Ottawa ON K1A 0N4
Canada

Bibliothèque nationale
du Canada

Acquisitions et
services bibliographiques

395, rue Wellington
Ottawa ON K1A 0N4
Canada

Your file Votre référence

Our file Notre référence

The author has granted a non-exclusive licence allowing the National Library of Canada to reproduce, loan, distribute or sell copies of this thesis in microform, paper or electronic formats.

The author retains ownership of the copyright in this thesis. Neither the thesis nor substantial extracts from it may be printed or otherwise reproduced without the author's permission.

L'auteur a accordé une licence non exclusive permettant à la Bibliothèque nationale du Canada de reproduire, prêter, distribuer ou vendre des copies de cette thèse sous la forme de microfiche/film, de reproduction sur papier ou sur format électronique.

L'auteur conserve la propriété du droit d'auteur qui protège cette thèse. Ni la thèse ni des extraits substantiels de celle-ci ne doivent être imprimés ou autrement reproduits sans son autorisation.

0-612-22151-2

Canada

UNIVERSITY OF ALBERTA

LIBRARY RELEASE FORM

NAME OF AUTHOR: Kui Huang

TITLE OF THESIS: Structural studies of the interactions
between serine proteinases and protein
inhibitors

DEGREE: Doctor of Philosophy

YEAR THIS DEGREE GRANTED: 1995

Permission is hereby granted to the University of Alberta Library to reproduce single copies of this thesis and to lend or sell such copies for private, scholarly or scientific research purposes only.

The author reserves all other publication and other rights in association with the copyright in the thesis, and except as hereinbefore provided neither the thesis nor any substantial portion thereof may be printed or otherwise reproduced in any material form whatever without the author's prior written permission.

Kui Huang
Kui Huang
8515, 112 Street
Edmonton, Alberta
T6G 1K7

Date: Sept 12, 1995

UNIVERSITY OF ALBERTA

FACULTY OF GRADUATE STUDIES AND RESEARCH

The undersigned certify that they have read, and recommend to the Faculty of Graduate Studies and Research for acceptance, a thesis entitled "Structural studies of the interactions between serine proteinases and protein inhibitors" submitted by Kui Huang in partial fulfillment of the requirements for the degree of Doctor of Philosophy.

Michael James

Dr. Michael N.G. James

A. Richard Morgan

Dr. A. Richard Morgan

L.B. Smillie

Dr. Lawrence B. Smillie

Karl R. Kopecky

Dr. Karl R. Kopecky

David Eisenberg

Dr. David Eisenberg

Date: Sept 7, 1995

To my parents and Frank

ABSTRACT

The crystal structure of the complex of porcine pancreatic elastase with a chymotrypsin/elastase inhibitor from *Ascaris suum* has been solved to 2.4 Å by the molecular replacement method. The C/E-1 inhibitor displays a novel folding motif, confirming that serine proteinase inhibitors produced by *Ascaris* are structurally distinct from any other known protein inhibitor families.

The crystal structures of the complexes of *Streptomyces griseus* proteinase B (SGPB) with three P₁ variants of turkey ovomucoid inhibitor third domain (OMTKY3), Leu¹⁸, Ala¹⁸ and Gly¹⁸, have been determined and refined to high resolution. These structures have revealed that water molecules are redistributed in adjusting to the size and shape of the P₁ side-chains in the S₁ binding pocket of SGPB. There are extensive water-mediated hydrogen bonds in the interfaces of the complexes. For a homologous series of aliphatic straight side-chains (0 to 5 methylene/methyl groups), Gly¹⁸, Ala¹⁸, Abu¹⁸, Ape¹⁸, Ahx¹⁸ and Ahp¹⁸ variants, the binding free energy is a linear function of the hydrophobic surface area buried in the interface of the corresponding complexes. The correlation constant gives a reliable estimate of the hydrophobic effect of 34.1 cal mol⁻¹ Å².

The ionizable P₁ residues Glu18I and Asp18I of OMTKY3 display unusually high pK_a values (8.7 and 9.4 respectively) when complexed with SGPB. The 1.8 Å crystal structures of the Glu¹⁸, Gln¹⁸, Asp¹⁸ and Asn¹⁸ variants of OMTKY3 have been determined in complex with SGPB at pH 6.5 and at pH 10.7 (Asn¹⁸ at pH 6.5 only). A potassium ion has been

identified in the pH 10.7 structures of the Glu¹⁸ and Asp¹⁸ variants, but not in any other structures. Apparently the K⁺ ion has been recruited to the S₁ pocket to balance the negative charge of the P₁ residue at pHs above its pK_a. This K⁺ ion forms a salt bridge with the carboxyl groups of the P₁ side-chains. The energy cost of burying the salt bridge is approximately 5.6 Kcal/mol, as estimated from the pH-induced decrease in the association constants.

ACKNOWLEDGMENT

I would like to thank my supervisor, Dr. Michael James for introducing me to the exciting world of protein crystallography. Mike's exceptional insights and great enthusiasm about sciences have always fascinated and inspired me. His superb supervision and constant support have made my six year stay a wonderful period of learning and doing science.

I am very grateful to the friendly and generous helps from people in James's lab. Anita Sielecki has given me tremendous helps in every aspect of life and work. I can not imagine myself going through those difficult periods without her motherly care and professional advice. My appreciation of her love and supports is beyond words. I wish to thank Natalie Strynadka for her friendship and helps. Natalie kindly helped me to get familiar with computer programs and the area detectors. She has been a source of great inspiration and encouragement, both scientifically and personally. I regard Natalie as a role-model for myself. I also thank Marie Fraser for help on running programs and many other aspects of crystallography. I admire her judgment on a lot of matters. Marc Allair, Stanley Moore and Ernst Bergmann have assisted in collecting several data sets used in the thesis. It is always very delightful and stimulating to work with Marc. I thank Randy Read and Masao Fujinaga for many helpful discussions; Ron Meleshko for solving computer problems and Mae Wylie for help in preparing manuscripts. I Have enjoyed working with fellow graduate students in the lab, Steve Mossiman, Kathy Bateman, Nina Khazanovich and Amir Khan. Amir has carefully made grammatical corrections in several parts of my thesis. Kathy and Steve have helped me to practice my defense. I also want to acknowledge the help from several people who worked previously in the James lab; Cathy McPhalen, Helen Blanchard, Bret Church and Kathy Jones.

The ovomucoid inhibitor third domain variants were kindly provided by Dr. Michael Laskowski Jr. and his colleagues in Purdue

University. The Ascaris inhibitor was kindly provided by Dr. Robert Peanasky and colleagues in the University of South Carolina. Their work has made this thesis possible. I would like to thank in particular Wuyuan Lu for his timely supplies of protein samples, ideas and encouragement.

I would like to thank members of the examining committee, Dr. David Eisenberg, Dr. Larry Smillie, Dr. Richard Morgan and Dr. Karl Kopecky, who made very good suggestions for the final revision of this thesis.

This research was supported by the Medical Research Council of Canada, through a grant to the MRC group in Protein Structure and Function in the University of Alberta. Dr. Michael James kindly provided me a research assistantship during the six years of my graduate study.

Finally, I must thank Frank Ye for his love and support over the past years. I feel very fortunate to have him in my life. The six years we spent together in Canada is the most wonderful and memorable period of time in my life. Our parents, sisters, and brothers, who live overseas, are always very supportive and always wish us happiness and success.

TABLE OF CONTENTS

	Page
Chapter 1: Introduction.....	1
<i>Serine proteinases</i>	<i>5</i>
<i>Chymotrypsin-like proteinases.....</i>	<i>6</i>
<i>Catalytic mechanism of serine proteinases.....</i>	<i>8</i>
<i>Substrate specificity</i>	<i>13</i>
<i>Canonical inhibitors of serine proteinases.....</i>	<i>15</i>
<i>Elastase and Ascaris inhibitors.....</i>	<i>20</i>
<i>Streptomyces griseus proteinase B.....</i>	<i>22</i>
<i>Turkey ovomucoid inhibitor third domain.....</i>	<i>23</i>
<i>Crystallographic Aspect.....</i>	<i>30</i>
REFERENCES.....	35
 Chapter 2: The Molecular Structure of the Complex of the Chymotrypsin/Elastase Inhibitor from <i>Ascaris lumbricoides</i> suum with Porcine Elastase	
INTRODUCTION.....	43
RESULTS AND DISCUSSION.....	46
<i>Overall structure of the C/E-1 inhibitor.....</i>	<i>46</i>
<i>Geometry of the reactive-site loop.....</i>	<i>56</i>
<i>Enzyme-inhibitor complex.....</i>	<i>62</i>
<i>Comparison of elastase in the free and complexed states.....</i>	<i>66</i>

Chemical determination of disulfide bond pairing.....	67
Biological implications.....	69
MATERIALS AND METHODS.....	71
Crystallization of the complex.....	71
Data collection.....	71
Structure determination.....	72
Refinement.....	74
REFERENCES	76

Chapter 3: Water Molecules Participate in Proteinase-Inhibitor Interactions: Crystal Structures of the Leu¹⁸, Ala¹⁸ and Gly¹⁸ Variants of Turkey Ovomucoid Inhibitor Third Domain Complexed with *Streptomyces Griseus* Proteinase B

INTRODUCTION	84
RESULTS.....	89
N-terminal truncation of OMTKY3 has no effect on the reactive-site conformation.....	89
Conformational changes of one substrate binding loop of SGPB.....	93
A phosphate binding site in the interaction region.....	95
Comparison of common water molecules in the complexes.....	97
Redistribution of water molecules in the S ₁ in the S ₁ binding pocket.....	99
Water molecules participate in the	

enzyme-inhibitor interaction.....	103
DISCUSSION	110
MATERIALS AND METHODS.....	118
Crystallization.....	118
Data collection.....	120
Structure determination and refinement.....	120
Structure comparison.....	122
Modeling of the non-coded amino acid variants of OMTKY3.....	122
Calculation of solvent accessible surface area.....	122
REFERENCES.....	124

**Chapter 4: Estimating the Energetic Effects of
Burying a Salt Bridge in a Hydrophobic
Pocket: Crystal Structures of the Glu¹⁸,
Gln¹⁸, Asp¹⁸ and Asn¹⁸ Variants of Turkey
Ovomucoid Third Domain in Complex with
Streptomyces Griseus Proteinase B at
Various pHs.**

INTRODUCTION.....	129
RESULTS.....	135
Quality of the structures of various complexes.....	135
Overall structures of the complexes.....	138
Structure of the Glu ¹⁸ variant at pH 6.5.....	142
Structure of the Glu ¹⁸ variant at pH 10.7.....	147
Structures of the Gln ¹⁸ variant at pH 6.5 and pH 10.7.....	152

Structure of the Asp ¹⁸ variant at pH 6.5.....	157
Structure of the Asp ¹⁸ variant at pH 10.7.....	161
Structure of the Asn ¹⁸ variant at pH 6.5.....	166
DISCUSSION.....	169
MATERIALS AND METHODS.....	174
Crystallization.....	174
Data collection.....	175
Structure determination and refinement.....	177
Structure comparisons.....	178
REFERENCES	179
Chapter 5: Conclusions.....	182

LIST OF TABLES

	Page
Table 1.1 The equilibrium constants for the association of 25 turkey ovomucoid P ₁ variants with six different serine proteinases.....	25
Table 2.1 Comparison of the P ₃ to P ₂ ' residues.....	60
Table 2.2 Interactions (within 4.0 Å) made between the C/E-1 inhibitor and PPE.....	64
Table 2.3 Data collection and structural refinement....	72
Table 3.1 Statistics of structural refinement.....	89
Table 3.2 The comparison of solvent structure in the three complexes.....	97
Table 3.3 Water molecules bound in the S ₁ binding pocket of SGPB.....	99
Table 3.4 The hydrogen-bonding and non-bonded interactions in the interface of the complexes of SGPB and OMKTY3.....	103
Table 3.5 The binding equilibrium constants and calculated buried surface areas.....	109
Table 3.6 Statistics of data collection.....	120
Table 4.1 Statistics of structural refinement.....	137
Table 4.2 RMS differences of main-chain atoms among the structures of the various complexes.....	141
Table 4.3 RMS differences of the bound water molecules.....	142

Table 4.4	Positional differences of solvent molecules bound in the S_1 pocket.....	144
Table 4.5	Water mediated hydrogen bonds and non-bonded contacts in the six complexes....	147
Table 4.6	Hydrogen-bonding interactions involving Asn33I and Asn36I in the six complexes.....	154
Table 4.7	Statistics of diffraction data.....	177

LIST OF FIGURES

	Page
Figure 1.1	Nomenclature of the substrate binding subsite....7
Figure 1.2	Classification of serine proteinase inhibitors.....14
Figure 1.3	Mechanism of inhibition.....16
Figure 2.1	Schematic representation of the C/E-1 inhibitor.....47
Figure 2.2a	An all-atom representation of the C/E-1 inhibitor.....48
Figure 2.2b	Sequence, secondary structure and disulfide bridges of the C/E-1 inhibitor.....48
Figure 2.3a	Hydrogen-bonding patterns in β -sheet 1.....49
Figure 2.3b	Schematic representation of the inter-strand hydrogen-bonds of β -sheet 1.....49
Figure 2.4a	Hydrogen-bonding pattern in β -sheet 2 and the small β -ladder.....51
Figure 2.4b	Schematic representation of the inter-strand hydrogen-bonds in β -sheet 2 and the β -ladder....51
Figure 2.5	Sequence alignment of the five homologous Ascaris C/E inhibitors with ATI.....52
Figure 2.6	The salt-bridge/hydrogen-bond network centered on Arg48I.....56
Figure 2.7	The salt-bridge/hydrogen-bond network centered on Arg44I.....57
Figure 2.8	Electron density in the active site of the complex.....58

Figure 2.9	Superposition of residues P_3 to P_2' of the reactive-site loop.....	59
Figure 2.10	A schematic representation of the penetration of Arg217A.....	62
Figure 3.1	Electron density map showing part of the active site of the complex of SGPB with the Ala ¹⁸ variant of OMTKY3.....	90
Figure 3.2	Superposition of the Leu ¹⁸ variant of OMTKY3 with native OMTKY3.....	91
Figure 3.3	Superposition of the uncomplexed SGPB with SGPB in complex with the Leu ¹⁸ variant.....	93
Figure 3.4	The phosphate binding site in the Leu ¹⁸ variant complex.....	95
Figure 3.5	Redistribution of water molecules bound in the S_1 binding pocket.....	100
Figure 3.6	Water-mediated hydrogen bonds in the complex of SGPB with the Leu ¹⁸ variant of OMTKY3.....	105
Figure 3.7	Water-mediated hydrogen bonds in the complex of SGPB with the Ala ¹⁸ variant of OMTKY3.....	107
Figure 3.8	Linear correlation of the binding free energy $-\Delta\Delta G$ with the hydrophobic surface area buried in the interface	111
Figure 4.1a	The electrostatic potentials on the van der Waals surface of SGPB.....	134
Figure 4.1b	The insertion of the P_1 side-chain into	

	the S ₁ binding pocket.....	134
Figure 4.2	Superposition of all seven complexes.....	140
Figure 4.3	A representative portion of the electron density map of the Glu ¹⁸ (N) structure.....	145
Figure 4.4	Hydrogen-bonding interactions in the Glu ¹⁸ (N) structure.....	146
Figure 4.5	The identification of a K ⁺ ion in the Glu ¹⁸ (H) structure.....	149
Figure 4.6	Superposition of the Glu ¹⁸ (H) and the Glu ¹⁸ (N) structures.....	152
Figure 4.7	A representative portion of electron density map of the Gln ¹⁸ (N) structure.....	156
Figure 4.8	Superposition of the Gln ¹⁸ (N) and the Glu ¹⁸ (N) structures.....	157
Figure 4.9	Electron density corresponding to the O ²¹ of Gln18 in the Gln ¹⁸ (N) structure.....	159
Figure 4.10	A representative portion of electron density map of the Asp ¹⁸ (N) structure.....	160
Figure 4.11	Hydrogen-bonding interactions in the Asp ¹⁸ (N) structure.....	161
Figure 4.12	The identification of a K ⁺ ion in the Asp ¹⁸ (H) structure.....	163
Figure 4.13	Superposition of the Asp ¹⁸ (H) and the Glu ¹⁸ (H).....	164
Figure 4.14	Superposition of the Asp ¹⁸ (H) and the Asp ¹⁸ (N) structures.....	166

Figure 4.15 A representative portion of the electron
density map of the Asn^{18} structure.....168

Figure 4.16 Superposition of the $\text{Asn}^{18}(\text{N})$ and the
 $\text{Asp}^{18}(\text{N})$ 169

LIST OF SYMBOLS AND ABBREVIATIONS

Å	Ångstrom (10^{-10} meters)
Abu	α -amino butyric acid
Ape	α -amino pentanoic acid (norvaline)
Ahp	α -amino heptanoic acid
Ahx	α -amino hexanoic acid (norleucine)
ATI	Trypsin inhibitor-1 from <i>Ascaris suum</i>
B	Thermal motion parameter
BPTI	Bovine pancreatic trypsin inhibitor (Kunitz)
BT	Bovine trypsin
CHYM	α -chymotrypsin
CMTI-I	Squash seed trypsin inhibitor-1
C/E-1	Chymotrypsin/elastase inhibitor-1 from <i>Ascaris suum</i>
CI-2	Barley chymotrypsin inhibitor 2
$ F_o $	Observed structure factor amplitude
$ F_{cal} $	Calculated structure factor amplitude
HLE	Human leukocyte elastase
Hse	Homoserine
K_a	Association equilibrium constant
K_a^{obs}	Observed association equilibrium constant

k_{cat}	First-order catalytic rate constant
K_m	Michaelis-Menten constant
MAD	Multiple anomalous dispersion
MIR	Multiple isomorphous replacement
NMR	Nuclear magnetic resonance
P_1	The residue contributing the carbonyl portion to the reactive-site peptide bond
OMJPQ3	Japanese quail ovomucoid inhibitor third domain
OMTKY3	Turkey ovomucoid inhibitor third domain
OMSVP3	Silver pheasant ovomucoid inhibitor third domain
PCI	Polypeptide chymotrypsin inhibitor from potato
PPE	Porcine pancreatic elastase
R	Standard crystallographic residual
rms	root-mean-square
S_1	The preformed cavity in the enzyme to which the P_1 residue binds
SGPA	<i>Streptomyces griseus</i> proteinase A
SGPB	<i>Streptomyces griseus</i> proteinase B
SGPE	<i>Streptomyces griseus</i> proteinase E
SLPI	Acid-stable human mucous proteinase inhibitor
SSI	<i>Streptomyces</i> Subtilisin inhibitor
STI	Soybean trypsin inhibitor

α_{cal}

Calculated structure factor phase

CHAPTER 1

INTRODUCTION

The limited proteolytic activity of proteinases is central to many physiological processes. In addition to their role in normal processes such as digestion, blood-coagulation, processing of protein precursors (zymogens and prehormones, etc.) and degradation of defective and exogenous proteins, proteinases are also involved in viral infection, apoptosis, tumor invasion and metastasis. Excessive proteolytic activity is implicated in many diseases such as arthritis, emphysema, hypertension, ulcer and pancreatitis. Therefore, the inhibition of proteolytic activity is sometimes more desirable than the proteolytic activity itself. Whereas microorganisms produce small non-proteinaceous inhibitors of proteinases, mammals always produce proteinaceous inhibitors to regulate and prevent unwanted proteolysis. It is estimated that about 10% of the proteins in plasma are proteinase inhibitors.

How inhibitors inactivate proteinases has been a focus of research for many chemists, biochemists and structural biologists, using a combination of kinetic, thermodynamic, site-directed mutagenesis and structural approaches. The investigations into the inhibition of proteinases by their inhibitors have led to the identification of the catalytic residues in various enzymes. The multiple approaches have led to an understanding of the catalytic mechanisms of proteinases, and the basis for enzymatic rate enhancement in general. The highly specific

interactions between proteinases and inhibitors, as demonstrated in these studies, have provided important information about the nature of molecular recognition and the free energy of association. In a more practical sense, structure-based drug design has greatly benefited from the structural studies of complexes between proteinases and various inhibitors.

Classically, proteinases are divided into four superfamilies according to the chemical nature of their catalytic residues. These four families are serine proteinases, cysteine proteinases, aspartic proteinases and metallo-proteinases, each utilizing a distinct catalytic apparatus (Reviewed by James, 1994). Serine proteinases, aspartate proteinases and metallo-proteinases have similar active site geometries despite sequence and structural dissimilarities, providing a classic example of convergent evolution for similar catalytic processes (James, 1994). Recently, a Thr residue has been implicated as a nucleophile in the proteasomes. It is likely that new proteinase families will emerge and modify the classical way of classification.

Proteinase inhibitors can be separated into three major classes. The first class are transition state analogs, of which the best studied inhibitors are peptide aldehydes, ketones, boronic acids and organophosphorus inhibitors. These inhibitors react with the nucleophile of a cognate enzyme, either reversibly or irreversibly, and form a non-productive transition-state adduct with the enzyme. The second class are ground-state substrate analogs, such as protein inhibitors and cyclic peptides. These substrate-like inhibitors form a very stable Michaelis complex with their cognate enzymes, and are cleaved at

extremely slow rates. The third class are mechanism-based, or suicidal inhibitors. The interactions of such an inhibitor with a cognate enzyme unmask a reactive group present in the inhibitor, which can then irreversibly inactivates the enzyme.

Protein inhibitors are ground-state substrate analogs, comprising thousands of naturally-occurring proteins of various lengths. They are produced by all kinds of organisms for various physiological purposes. Laskowski and Kato first introduced the rational nomenclature to classify protein inhibitors into separate families, on the basis of their sequence homology, disulfide bond topology, target proteinases and inhibition mechanism (Laskowski & Kato, 1980). The majority of protein inhibitors characterized so far are directed toward serine proteinases.

The main theme of this thesis is the study of how protein inhibitors interact with serine proteinases. A total of 11 crystal structures of 8 different complexes between a serine proteinase and a protein inhibitor have been determined and analyzed herein. The "Introduction" tends to provide a general survey of serine proteinases and protein inhibitors. For more exhaustive coverage, several excellent reviews are recommended below:

Books:

Proteinase inhibitors

Barrett AJ and Salvensen G, eds, Elsevier, Amsterdam, 1986.

Proteolysis and protein turnover

Bond JS and Barrett AT, eds, Brookfield, VT: Portland press, 1994.

Review papers:

On proteinases: Perona and Craik, 1995.

On proteinase inhibitors: Laskowski and Kato, 1980; Bode and Huber, 1992.

Serine proteinases

Serine proteinases are among the best studied enzymes, both enzymatically and structurally. Serine proteinases are characterized by an unusually reactive serine residue that is essential for its activity. A conserved histidine and an aspartate are also critical for catalysis. The essential role of the serine was established from the observation that diisopropyl fluorophosphate inhibits chymotrypsin by esterifying the Ser (Hartley, 1960). The importance of a histidine was shown by the use of tosyl-phenylalanyl-chloromethane (Schoelman & Shaw, 1963), which initially reacts with and forms a tetrahedral intermediate with the catalytic Ser, and subsequently alkylates the catalytic His.

Classically, serine proteinases were divided into two superfamilies, the chymotrypsin family and the subtilisin family (Hartley, 1970; James, 1976). The crystal structures of bovine chymotrypsin (Matthews et al., 1967) and of subtilisin (Wright et al., 1969), the prototypes of the two families, were among the first crystal structures of enzymes. It was found that the two enzymes displayed completely different tertiary structures. Chymotrypsin adopts a bilobal fold, consisting of two β -barrel domains. Subtilisin has a single-domain structure, with α -helices packed against a central β -sheet. In both enzymes, the catalytically important residues, Ser, His and Asp are clustered in a surface crevice of the enzyme, and the spatial arrangements of these residues in the two enzymes are strikingly similar. In addition, both enzymes have an oxyanion binding site called

the "oxyanion hole", which comprises two nitrogen atoms. As an element of the catalytic apparatus, the oxyanion hole is essential for stabilizing the reaction intermediates.

With the ever increasing discovery of new serine proteinases, it has been realized recently that proteins are not limited to the chymotrypsin-fold or subtilisin-fold in order to achieve proteolytic activity. The crystal structure of wheat serine carboxypeptidase II (Liao et al., 1990) revealed conservation of the essential features of the catalytic apparatus within a third distinct protein fold - an α/β hydrolase fold. Furthermore, it has been proposed that the immunoglobulin fold might represent a fourth structural family of serine proteinases. The crystal structure of an antibody capable of catalyzing the hydrolysis of diesters revealed the presence of a Ser-His catalytic dyad structurally similar to other serine proteinases (Zhou et al., 1994). The similar design of the active site geometries displayed by the above four different structural families of serine proteinases represents an excellent example of convergent enzyme evolution, suggesting that the design is essential for serine proteinases to achieve high proteolytic efficiency (high k_{cat}) and specificity (high k_{cat}/K_m). It is very likely that new serine proteinases with novel folds will be discovered which assume the characteristic active site geometry of the known serine proteinases.

Chymotrypsin-like serine proteinases

There is a wealth of structural information about chymotrypsin-like serine proteinases. Crystal structures are available for bovine chymotrypsin (Matthews et al., 1967), bovine and rat trypsins (Ruhlmann

et al., 1973; Sprang et al., 1987), porcine pancreatic and human leukocyte elastases (Watson et al., 1970; Navia et al., 1989), thrombin (Bode, 1989a), factor Xa (Padmanabhan et al., 1993), complement factor D (Narajana et al., 1994), kalikrein (Bode et al., 1983), rat tonin (Fujinaga & James, 1987) and rat mast cell protease II (Remington et al., 1988). In addition, structures have been determined for five bacterial serine proteinases: *Streptomyces griseus* trypsin and *Streptomyces griseus* proteinase A, B, E (SGT, Read & James, 1988; SGPA, Delbaere et al., 1975; SGPB, Delbaere et al., 1979; SGPE, Nienaber et al., 1993) and the *Lysobacter enzymogenes* α -lytic protease (Brayer et al., 1979).

Members of the chymotrypsin family display homologous sequences and a similar tertiary structure that is typified by bovine chymotrypsin. These enzymes possess a bilobal β -barrel structure with the active site cleft located in the junction. The substrate polypeptide spans across the cleft, with the catalytic Ser and His overlooking the scissile bond. On either side (the C-terminal domain) of the catalytic residues are a set of specificity pockets, that are preformed cavities to accommodate the side-chains of the bound peptide segment. By convention, the nomenclature of Berger and Schechter (Berger and Schechter, 1970) is used to identify the substrate specificity pockets (Figure 1.1). They are numbered S_1 , S_2 etc., going away from the catalytic site towards the N-terminus of a bound peptide substrate and S_1' , S_2' , etc., towards the C-terminus of the peptide. The substrate residues are correspondingly numbered P_1 , P_2 and P_1' , P_2' with the cleavage bond (scissile bond) between P_1 and P_1' . A good substrate is complementary to the active

site cleft, so that its sidechains can fit into the respective specificity pockets without causing any steric collisions.

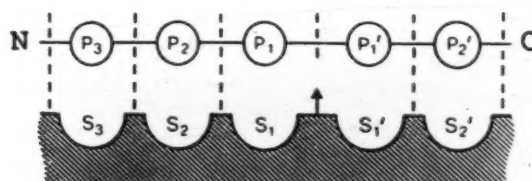


Figure 1.1 Nomenclature of the substrate binding subsites.

Catalytic mechanism of serine proteinases

On the basis of the conserved active site geometry, it was proposed (and subsequently confirmed experimentally) that all serine proteinases act by an identical mechanism. Serine proteinases cleave peptide bonds by forming a covalently attached tetrahedral transition intermediate with the substrate. In the proposed mechanism (using the example of chymotrypsin), Ser195 acts as a nucleophile, and His57 as a general base in removing the proton from Ser195 O^V. The oxyanion hole stabilizes the developing negative charge on the carbonyl oxygen atom of the scissile bond. The reaction involves the formation of a Michaelis complex, followed by a tetrahedral intermediate, an acyl-enzyme intermediate, a second tetrahedral intermediate with the attack of a water nucleophile and finally a deacylation step.

The reaction is initiated by the formation of a Michaelis complex between the enzyme and the bound substrate. The substrate forms a five-residue antiparallel β -sheet with the enzyme, through backbone hydrogen

bonds, mostly on the N-terminal side of the scissile bond. The complex is further stabilized by inserting the substrate sidechains into the respective specificity pockets. These interactions serve to orient the carbonyl carbon in a position that is ideal for nucleophilic attack by Ser195 O^γ. The carbonyl oxygen forms hydrogen bonds with the nitrogen atoms constituting the oxyanion hole.

The reaction proceeds to nucleophilic attack of the carbonyl carbon by the Ser195 O^γ. The nucleophilicity of Ser195 O^γ is thought to be enhanced by the hydrogen bond to His57 N^{ε2}, as well as by the partial positive charge of the carbonyl carbon. The polarization of the C=O bond is enhanced by positioning of the carbonyl oxygen in the oxyanion hole, where it forms hydrogen bonds with backbone nitrogen atoms of Ser195 and Gly193. Upon nucleophilic attack, the carbonyl carbon undergoes rehybridization from sp² to sp³, resulting in a tetrahedral intermediate that is covalently linked to Ser195 O^γ. The C=O double bond becomes a single bond, so that the oxyanion forms stronger(shorter) hydrogen bonds within the oxyanion hole. The proton of Ser195 O^γ is transferred to His57 N^{ε2}, and the resulting imidazole ion is stabilized by forming a hydrogen bonded ion pair with Asp102. Thus the first tetrahedral intermediate of the reaction is formed.

The proton on His57 N^{ε2} is then transferred to the leaving group, causing the cleavage of the scissile bond. An acyl-enzyme intermediate is formed while the first product is released from the active site. It is believed that the carbon maintains a partly pyramidal configuration, and carries a partial positive charge (James et al., 1980). The partial

negative charge on the carbonyl oxygen is stabilized by interactions with the oxyanion hole.

The deacylation step is almost the reverse of the acylation process, except that the nucleophile is a water molecule rather than the Ser195 O⁷. An incoming deacylation water approaches the carbonyl carbon from the just-vacated leaving group side. This water serves as the nucleophile in the deacylation reaction. The nucleophilicity of the water is proposed to be induced by the partial positive charge of the carbonyl carbon and the hydrogen bond to His57 N⁴². This water molecule transfers one of its protons to His57 N⁴² and the resulting strong nucleophile OH⁻ attacks the carbonyl carbon. A second tetrahedral intermediate is formed, with the negative charge which develops on the carbonyl oxygen again stabilized by the oxyanion hole interactions. Finally, the His57 N⁴² transfers the proton back to Ser195 O⁷, and the second product is released.

The proposed mechanism is supported by substantial structural evidence. Some of the reaction intermediates have been observed crystallographically. The crystal structures of the complexes between serine proteinase and the canonical protein inhibitors resemble the Michaelis complex. Structures of bound transition state analogs shed light on why a tetrahedral intermediate binds a cognate enzyme more tightly than a ground-state substrate (James et al., 1980; Matthews et al., 1975; Poulos et al., 1976; Stroud et al., 1974). Acyl-enzyme adducts have been observed crystallographically in γ -chymotrypsin and SGPA (Dixon et al., 1991; Harel et al., 1991; Blanchard et al., 1994). The enzyme-product complex observed in a crystal structure of SGPA and a

tetrapeptide aldehyde mimics the second tetrahedral intermediate in the deacylation step, but the distance from Ser195 O^γ to the carbonyl carbon (2.6 Å) is not a covalent bond (James et al., 1980).

Mutational analyses of both subtilisin and trypsin have confirmed the essential role of Ser195 and His57 in providing rate acceleration. Replacement of the catalytic Ser221 and His64 residues of subtilisin with alanine results in decreases of 10^4 to 10^6 fold in k_{cat} (Carter & Wells, 1987). Similar results were obtained by analogous mutations of Ser195 and His57 in rat trypsin (Corey & Craik, 1992). The catalytic triad imparts a total rate enhancement of a factor of $\sim 10^6$ over the uncatalyzed reaction (Carter & Wells, 1987).

The essential role of the oxyanion hole should also be emphasized here. There is considerable residual activity in the absence of the catalytic triad ($k_{cat}/k_{uncat} = 10^3$) (Carter & Wells, 1987). This activity has been attributed to the remaining binding determinants, particularly the oxyanion hole, that stabilize the transition state complex.

The involvement of Asp102 in the catalysis has been verified by the Asp102-Asn mutation in rat trypsin, where 100 fold reduction was observed (Craik et al., 1987). However, it is noticed that when chymotrypsin, subtilisin and wheat carboxypeptidase are superposed, the aspartate residues are in different positions with respect to the Ser-His pair and the hydrogen bonds between the Asp and the His are not the same in the three enzymes. It is then proposed that the "catalytic triad" should be regarded as two diads, a His-Asp diad and a His-Ser diad, and the relative orientation of one diad with respect to the other is not significantly important (Liao, 1992).

The position of the deacylating water molecule is debatable. The time-resolved Laue crystallography on trypsin and a slowly hydrolyzed substrate has suggested that the water molecule approaches the His57 N⁴² from the amino terminal side of the carbonyl carbon, and is located 3.7 Å from the carbonyl carbon (Singer et al., 1993). As opposed to this, Perona et al. have suggested that the deacylating water is expected to approach the His 57 N⁴² from the leaving group side, which is geometrically favorable (Perona et al., 1993). The deacylating water they have suggested, which is in a conserved site in five crystal structures of peptide transition state analog inhibitor complexes with human leukocyte elastase, trypsin and α -lytic protease, sits on the leaving group's side and at a distance of 3.5-3.7 Å from the carbonyl carbon. Two additional positions have also been suggested from the studies of acyl-enzyme adducts of γ -chymotrypsin (Dixon et al., 1991) and SGPA (Blanchard & James, 1994); both occur at the leaving group side of the inhibitor, but at different distances and angles to the carbonyl carbon.

Substrate specificity

Serine proteinases of the chymotrypsin family display a diverse substrate specificity. The primary specificity determinant is the chemical nature and the volume of the S₁ pocket, which accommodates the P₁ residue. Enzymes displaying trypsin-like specificity have an Asp at the bottom of the S₁ pocket, thereby preferring Arg or Lys at the P₁ position. A histidine triad identified in the bottom of the S₁ pocket of the bacterial serine proteinase *Streptomyces griseus* E has been found to modulate specificity towards P₁ Glu. Enzymes exhibiting

chymotrypsin-like specificity possess a large hydrophobic pocket, thus large hydrophobic side-chains are favored. The S_1 pocket in elastase-like proteinases is relatively small, so that small aliphatic P_1 residues are favored. Other than the charged or the hydrophobic residues, polar residues such as Gln and Asn, are not particularly favored or rejected by any proteinases. Usually proteinases are very discriminating, showing strict specificity due to the rigid nature of the S_1 pocket. Interestingly, a collagenase from crab is highly active towards hydrophobic residues and Arg (Tsu et al., 1994). The underlying molecular recognition mechanism has yet to be elucidated by crystal structure studies.

Protein engineering aimed at modifying the activity and specificity of proteinases have greatly extended our understanding about the structure-function relationship of these proteins. Firstly, it is found that specificity and rate enhancement are closely related, as mutants designed to change an enzyme's specificity also changed the rate of catalysis and vice versa (Hedstrom et al., 1992). It has been further demonstrated that the specificity of an enzyme is expressed in the chemical steps of the reaction rather than in ground-state substrate binding (Hedstrom et al., 1994). Secondly, it has been realized that modification of the S_1 pocket is not the only component necessary and sufficient to change the enzyme's specificity. Regions that have no direct contact with the substrate also have great impact on the specificity (Hedstrom et al., 1994).

Canonical inhibitors of serine proteinases

Figure 1.2 is a nearly-complete list of the protein inhibitor families of serine proteinase that have been characterized so far, and at least 13 of these have one or more crystal structures determined to date. Most of these inhibitors follow a standard mechanism proposed by Laskowski and Kato, and are thus termed as "canonical inhibitors" (Laskowski & Kato, 1980). Some inhibitors, such as the serpins and the hirudins are known to act by different mechanisms.

According to the standard mechanism, canonical inhibitors bind to a cognate enzyme in the same manner as a good substrate, but are cleaved at an extremely slow rate. Structural information is available for a number of canonical inhibitors from various families, in free form or in complex with their cognate enzymes. These inhibitors are small proteins with from 20 to 190 amino acid residues. Inhibitors from different families display unrelated sequences and structures. A relatively comprehensive summary of the various folds of the inhibitors is given by Bode & Huber (1992). The hallmark of a canonical inhibitor is an exposed binding loop of a characteristic canonical conformation. The binding loop is complementary to the active site cleft of a target enzyme, and an extremely tight Michaelis complex is formed upon the binding (Reviewed by Read & James, 1986; Bode & Huber, 1992).

Kinetic studies have shown that the specificity constant of a canonical inhibitor k_{cat}/K_m is comparable to that of a good substrate. However, the individual values of the k_{cat} and K_m are extremely small relative to that of a good substrate. The inhibitor binds to a cognate enzyme much better than does a good substrate for the following reasons:

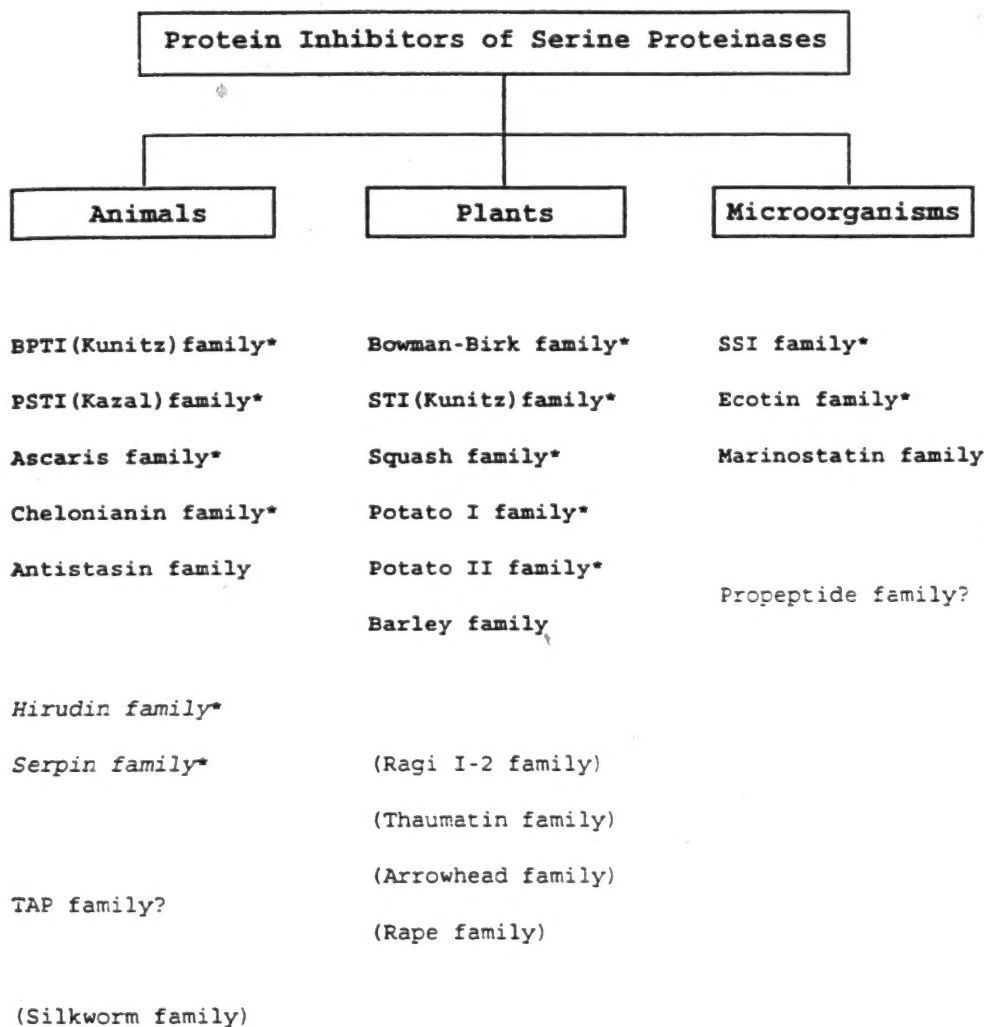


Figure 1.2 Classification of Serine Proteinase inhibitors. Canonical inhibitors are in bold; families that are known not to obey the standard mechanism are italicized; provisional families are in parentheses; question-marked are families whose mechanism is in doubt; families followed by "*" have X-ray crystal structures available.

1. The favorable entropic component to binding free energy. The reactive loop of the inhibitor is relatively rigid, partly due to extensive interactions with the protein core. There is little conformational entropy loss upon binding to the enzyme. In contrast, substrates are much more flexible, capable of adopting many conformations in solution but only one or a few in the Michaelis complex. Thus, there is a large unfavorable entropic component to the binding free energy of a substrate.

2. The enthalpic contribution. There are extensive interactions between the bound inhibitor and the enzyme. The contact surface area is large, involving about a dozen residues from the inhibitor and more from the enzyme. The contact area is significantly larger in inhibitor-proteinase complexes than that in substrate-proteinase complexes. For example, the contact area between turkey ovomucoid inhibitor third domain (OMTKY3) and *Streptomyces griseus* proteinase B (SGPB) is about 1300 Å², as opposed to 1100 Å² between SGPB and an octapeptide substrate with the sequence of the P₅ to P₃' residues of OMTKY3. The highly complementary interface is closely packed, with residual space filled by bound water molecules which introduce indirect interactions between the enzyme and the inhibitor. An antiparallel β-sheet is formed between the inhibitor and the enzyme, characterized by four strong hydrogen bonds. The dipole of the C=O bond is stabilized by two factors: the oxyanion hole and the close proximity of C to Ser195 O^γ (2.7 Å). In addition, the favorable interactions in the specificity pockets, particularly the P₁-S₁ interactions, further stabilize the complex significantly.

The exceptionally low k_{cat} of canonical inhibitors is closely related to the low K_m . As depicted in Figure 1.3, the inhibitor experiences a higher activation energy for hydrolysis than that of a substrate. The exceptionally tight Michaelis complex forms an energy sink, which is equivalent to increasing the activation energy toward the tetrahedral transition state. To undergo the deformation (from trigonal

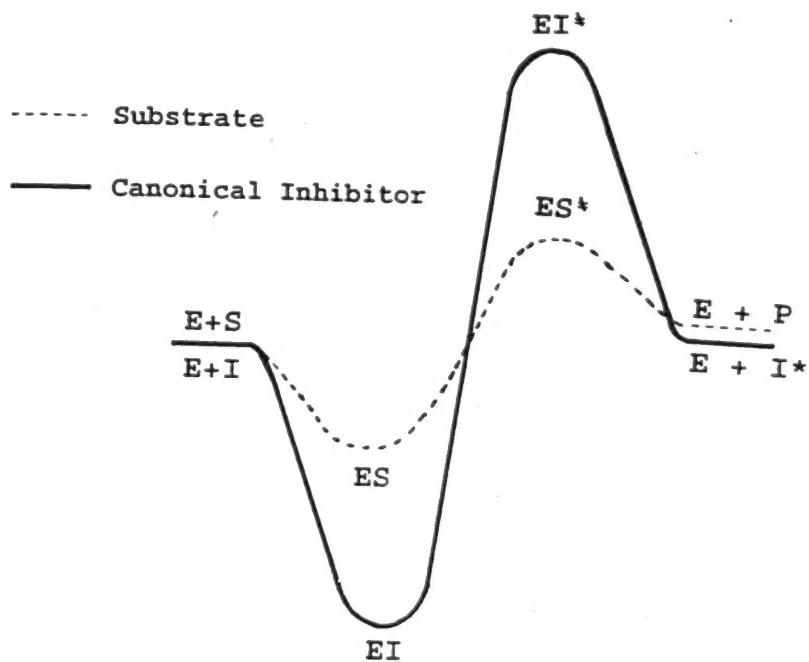


Figure 1.3 Mechanism of inhibition

planar carbonyl carbon atom to a tetrahedral carbon, i.e., rehybridization from sp^2 to sp^3) that is necessary for forming the transition state intermediate, some hydrogen bonds have to be disrupted. Furthermore, the reactive-site loop is very rigid due to multiple interactions with the enzyme and with the core structure of the inhibitor. Steric strains will have to be introduced to the inhibitor to achieve the tetrahedral transition state, thus elevating the energy level of the transition state intermediate. A high activation energy arises from the combination of the low energy Michaelis complex and the high energy tetrahedral intermediate.

A second factor is that there are extensive interactions at the P' sites, which make it difficult for the leaving group to leave even if the scissile bond can be cleaved. The retention of the leaving group is equivalent to increasing the concentration of product, thus shifting the reaction towards the substrate side. Laskowski and colleagues have shown that the scissile-bond cleaved inhibitors of OMTKY3 associate with the cognate proteinases to form the same stable complex as the virgin species, and utilized this property to synthesize OMTKY3 variants from two peptide fragments (Aldelt & Laskowski, 1983; Bigler et al., 1993).

Canonical inhibitors have great potential in guiding structure-based drug design, because substrate-like inhibitors display much higher specificity toward their cognate enzymes than do transition state analogs.

Elastase and Ascaris inhibitor

Porcine pancreatic elastase consists of 240 amino acids. Its physiological function is to digest protein in intestines. Excessive and premature activity of pancreatic elastase is linked to pancreatitis. It shares 40% sequence homology with human leukocyte elastase. Human leukocyte elastase has been associated with a number of inflammatory disorders including emphysema and other degenerative lung diseases (reviewed by Bode et al., 1989).

With regards to substrate specificity, elastase prefers small aliphatic residues, such as Ala and norvaline (Harper et al., 1984). The S_1 pocket is small due to the presence of two bulky residues, Val216 and Thr226 at the edge of the pocket.

Elastase has been the subject of extensive structural studies. More than 20 crystal structures of elastase, in free state and in complex with various small inhibitors, have been determined (reviewed by Bode et al., 1989). The structural studies have been facilitated by the fact that porcine pancreatic elastase crystallizes in a crystal form such that its active site was accessible to small inhibitors (up to 6 amino acid residues or their equivalent) when soaked with the crystals. It was also aided by the fact that elastase has a relatively rigid active site and there is little conformational change upon the inhibitor binding. Introduction of inhibitors to the native enzyme does not disrupt crystal contacts. It is particularly interesting that three homologous inhibitors of elastase show surprisingly different binding modes with the enzyme (Mattos et al., 1994). It is worthy mentioning that a non-peptide inhibitor against human leukocyte elastase has been

developed on the basis of porcine pancreatic elastase structure, and showing sufficient oral bioavailability to warrant further studies (Brown et al., 1994).

The *Ascaris* inhibitor used in this study is isolated from *Ascaris suum* (Peanasky et al., 1984). *Ascaris suum* is a parasitic worm that infects hog intestines. The worm produces a variety of protein inhibitors that are directed to the host's digestive enzymes (reviewed by Peanasky et al., 1987). The obvious physiological role of these inhibitors is to protect the worms from being degraded by host proteinases. It is interesting that the worm does not secrete the inhibitors, but rather takes up the host proteinases to its intestine where they are inactivated (Martzen et al., 1985). Understandably this is the most efficient way to protect itself, given the small weight of the worm and the large amount of digestive enzymes present in the intestines.

Among the protein inhibitors that *Ascaris suum* produces, five are found to inactivate chymotrypsin and elastase (Peanasky et al., 1984a), and three inactivate trypsin (Goodman & Peanasky, 1982). These inhibitors of 62 to 65 residues share 30% to 40% sequence identity (Babin et al., 1984). In the absence of any sequence homology to other known protein inhibitors, the *Ascaris* inhibitors represent a distinct inhibitor family called the "*Ascaris* inhibitor family". It has been demonstrated that the inhibitors follow the standard mechanism of inhibition (Peanasky et al., 1984b).

Chapter 2 reports the structure of the complex between porcine pancreatic elastase and a 63-residue inhibitor from *Ascaris suum*. The

structure was determined at 2.4 Å resolution, by the molecular replacement method. The inhibitor exhibits a novel folding motif characterized by two β -sheets and five disulfide bridges. The structure has revealed the correct linkage of the five disulfide bridges, and reassured the previous prediction that the *Ascaris* inhibitors directed towards serine proteinases indeed belong to a distinct family. Curiously, the inhibitor does not have a hydrophobic core. The overall tertiary structure is stabilized mainly by five disulfide bridges, hydrogen bonds and electrostatic interactions. It is intriguing that, without the assistance of a hydrophobic core, the inhibitor can assume correct linkage of the five disulfide bonds and the final fold.

Prior to this crystal structure, there had been no crystal structures of members of the *Ascaris* family. An NMR structure was done on a closely related inhibitor directed against trypsin (Gronenborn et al., 1990). The reported structure of the complex is the first crystal structure of pancreatic elastase inhibited with a large, protein inhibitor. The only other large inhibitor-elastase complex, the complex of human leukocyte elastase and ovomucoid inhibitor third domain, has been solved to 1.8 Å resolution (Bode et al., 1986); but the atomic coordinates have not been made available, presumably because of commercial interest.

Streptomyces Griseus proteinase B

SGPB is a serine proteinase isolated from pronase, the extracellular culture filtrate of *Streptomyces griseus*. The amino acid sequence was determined in the laboratory of Dr. L.B.Smillie (Jurasek et al., 1974). The X-ray structure of native SGPB was solved previously in

the laboratory of Dr. James (Coddington et al., 1974; Delbaere et al., 1979). SGPB displays low sequence identity with chymotrypsin (~20%) and is highly homologous to another bacterial serine proteinase, *Streptomyces griseus* A (62% identity). SGPB also displays a chymotrypsin-like specificity, preferring bulky hydrophobic P₁ residues such as Leu, Met, Phe and Tyr.

Crystal structures are known for free SGPB at pH 4.5 (orthorhombic crystal form, Delbaere et al., 1979) and at pH 7.0, the optimal pH for enzymatic activity (cubic crystal form, Blanchard & James, 1995). In addition, crystal structures are available for complexes of SGPB with OMTKY3 (Read et al., 1983) and with a chymotrypsin inhibitor from potato tubers (Greenblatt et al., 1989). The recombinant OMTKY3 variants that will be reported in this thesis are different from the previously determined OMTKY3 in that the first five residue of the N-terminus have been truncated to facilitate expression and purification. Measurements of association constants indicate that the truncation has no effect on the reactivity of the inhibitor (Wieczorek, et al., 1987).

Turkey ovomucoid inhibitor third domain

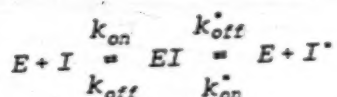
OMTKY3 is a paradigm of canonical inhibitors of serine proteinases. It is one of the 153 types of ovomucoid inhibitors from various species of birds that Laskowski and colleagues have isolated, sequenced and characterized over a period of 20 years. Avian ovomucoid inhibitors usually comprise three homologous domains. OMTKY3 is the third domain comprising 56 amino acids and three disulfide bonds, with Leu18 at the P₁ position. The wild-type OMTKY3 is a potent inhibitor of chymotrypsin, subtilisin, pancreatic elastase, human leukocyte elastase,

SGPB and SGPA (Bigler et al., 1993). The specificity of the inhibitor is primarily determined by the P₁ residue. OMTKY3 can be directed to inhibit trypsin when the P₁ Leu18 is mutated to Lys (Lu, 1994) and inhibits SGPE when the P₁ residue is mutated to Glu (Komiya, et al., 1991). A triple variant developed on the scaffolding of OMTKY3, Arg¹⁵-Lys¹⁷-Arg¹⁸, becomes a strong inhibitor of furin, a subtilisin-like prohormone convertase which prefers basic residues at the P₁ and P₁' positions (Lu et al., 1993). It has been shown that the reactive-site loop of OMTKY3 has certain flexibility, so that it can adjust to the small difference in structure in the substrate binding cleft (Fujinaga et al., 1987).

The idea of establishing a "sequence-reactivity algorithm" has motivated a collaborative study involving three groups: Dr. S. Anderson at Rutgers University, Dr. M. Laskowski Jr. at Purdue University, and Dr. M. James at the University of Alberta. The gene encoding a recombinant OMTKY3 (residues 6 to 56) has been cloned in the laboratory of Dr. S. Anderson, where the 20 P₁ variants have been generated with each of the 20 coded amino acids introduced to the P₁ position. In addition, five non-coded amino acid P₁ variants, Abu, Ape (norvaline), Ahx (norleucine), Ahp and homoserine, were produced by enzymatically linking two peptide fragments (Wieczorek & Laskowski, 1983; Wieczorek, et al., 1987; Bigler, et al., 1993). Of the two fragments, the N-terminal 13-amino acid peptide was chemically synthesized, with the non-coded amino acid at the C-terminus; whereas the C-terminal fragment was isolated from the scissile-bond cleaved OMTKY3 (with SGPB, at pH 1.5). The two fragments were cross-linked by the three disulfide bonds. Correct disulfide linkages yield a modified inhibitor in which the

scissile bond was cleaved. The modified inhibitor was then incubated with proteinase K to complete the closure of the scissile bond, and an intact ONTKY3 was generated.

The reactivity of each variant (represented by K_a), was tested in Dr. Laskowski's Lab at Purdue University. For a canonical inhibitor interacting with a cognate proteinase, ideally



K_a (association constant) and K_a^{obs} (observed association constant) can be defined as

$$K_a = \frac{[EI]}{[E] \times [I]} = \frac{[E_0] - [E]}{[E] \times ([I_0] - [E_0] + [E])}$$

$$K_a^{obs} = \frac{[EI]}{[E] \times ([I] + [I'])} = \frac{K_a}{1 + \frac{[I']}{[I]}}$$

where $[E]$, $[I]$, $[I']$ and $[EI]$ are the equilibrium concentrations of the enzyme, the virgin inhibitor, the modified inhibitor and the enzyme-inhibitor complex, respectively; $[E_0]$ and $[I_0]$ represent the total concentrations of the enzyme and the inhibitor. For the vast majority of canonical inhibitors, the hydrolysis occurs extremely slowly at neutral pH. During the period of incubation with the enzyme (typically from a few minutes to a few days), $[I']$ is virtually undetectable. Hence $K_a^{obs} = K_a$ when $[I'] = 0$.

After enzyme and inhibitor at known concentrations are mixed and the equilibrium is established, the free enzyme concentration $[E]$ can be determined by measuring the remaining proteolytic activity. The value of K_s can be obtained by measuring $[E]$ at various $[I_0]$ concentrations. The current technique developed in Dr. Laskowski's Lab is able to measure binding constants from 500 to 10^{13} M^{-1} , with an estimated error of $\pm 20\%$ (Bigler et al., 1993; Lu, 1994).

The equilibrium association constants have been measured for all the 25 P_1 variants interacting with six well-studied serine proteinases, chymotrypsin, subtilisin, SGPB, SGPA, pancreatic elastase and human leukocyte elastase. The thermodynamic data are summarized in Table 1.1 (Bigler et al., 1993; Lu, 1994). These thermodynamic measurements have systematically characterized the preference of the S_1 pocket in the six enzymes. The K_s for the Gly¹⁸ variant provides an indication as to how much binding free energy is attributable to the scaffolding of the inhibitor. When OMTKY3 interacts with SGPB, it has been shown that aliphatic hydrophobic P_1 residues provide additional stabilization to the complex. The K_s increases as the number of side-chain carbon atoms increases, and reaches the maximum at Ahp. In contrast, positively-charged, negatively charged and β -branched residues (Val, Ile and Thr) show destabilizing effects on the binding. The S_1 pocket of SGPB is sulfur-philic, since replacement of a carbon by a sulfur (Abu to Cys, Ahx to Met) elevates the K_s . Polar residues, such as Ser, Thr, Asn and Gln binds less tightly than comparable aliphatic residues.

These thermodynamic measurements comprise one of the most thorough investigations into enzyme-inhibitor interactions. Crystal structures

Table 1.1

The equilibrium constants for the association of 25 turkey ovomucoid P₁ variants with six different serine proteinases

	CHYM	PPE	SUBT	SGPA	SGPB	HLE
0 GLY	6.6E+06	9.0E+08	6.4E+08	4.5E+07	1.2E+07	2.2E+07
1 ALA	6.1E+07	4.2E+10	2.0E+10	1.9E+09	3.6E+08	1.0E+09
2 ABU	1.1E+09	<u>3.3E+11</u>	1.0E+11	1.0E+10	2.3E+09	1.4E+10
SER	4.2E+07	8.6E+08	2.4E+09	1.6E+08	5.0E+07	3.3E+07
CYS	2.4E+09	2.5E+10	<u>3.3E+11</u>	3.3E+11	9.8E+09	7.1E+09
3 APE	2.2E+10	2.4E+11	1.4E+11	8.3E+10	1.2E+10	1.0E+10
VAL	1.4E+08	8.8E+09	8.9E+08	2.1E+09	3.3E+08	1.4E+10
HSE	2.5E+09	7.5E+09	5.6E+10	6.5E+09	1.7E+09	2.6E+08
THR	9.8E+07	2.9E+10	2.6E+10	2.0E+09	2.5E+08	1.3E+09
PRO	<u>6.7E+04</u>	5.9E+05	<u>7.9E+04</u>	<u>4.6E+04</u>	<u>3.6E+04</u>	1.9E+05
4 AHX	8.0E+10	1.9E+11	6.7E+10	2.3E+11	2.1E+10	9.9E+08
LEU	1.9E+11	4.2E+10	3.4E+10	3.0E+11	<u>5.6E+10</u>	6.1E+09
ILE	9.9E+07	5.9E+09	1.4E+08	1.5E+08	2.9E+07	<u>2.1E+10</u>
MET	1.0E+11	1.3E+10	8.1E+10	2.6E+11	2.7E+10	7.3E+08
ASN	6.6E+08	6.9E+07	2.7E+09	2.6E+08	1.8E+08	9.0E+05
ASP	1.0E+06	7.3E+04	1.6E+07	5.1E+06	3.9E+06	<u>3.2E+03</u>
5 AHP	3.3E+11	2.4E+08	8.9E+10	<u>4.7E+11</u>	2.8E+10	1.3E+08
LYS	1.0E+08	4.6E+04	1.2E+08	5.6E+07	2.6E+08	3.8E+05
GLN	1.3E+09	4.3E+07	1.3E+10	1.2E+09	6.9E+08	2.3E+07
GLU	2.5E+06	8.4E+04	8.6E+08	7.9E+06	2.3E+06	4.1E+04
6 HIS	1.2E+09	3.7E+04	1.3E+10	2.6E+09	3.1E+09	1.1E+05
7 PHE	2.4E+12	4.9E+04	1.2E+10	1.8E+11	5.5E+09	9.9E+05
ARG	2.2E+08	<u>4.8E+03</u>	2.4E+08	8.2E+07	1.9E+08	3.3E+04
8 TYR	<u>8.1E+12</u>	7.5E+03	1.6E+10	1.0E+11	3.3E+09	8.3E+04
10 TRP	3.3E+12	2.5E+04	2.6E+09	8.9E+09	2.4E+09	1.8E+04

CHYM: α -chymotrypsin; PPE: porcine pancreatic elastase; SUBT, subtilisin Carlsberg; SGPA and SGPB: *Streptomyces griseus* proteinase A and B; HLE: human leukocyte elastase. The largest K_a in each column is boxed; the smallest K_a in each column is underlined.

have shown that enzyme-inhibitor interfaces are as densely packed as the residues in the interior of a protein. Therefore, factors stabilizing or destabilizing a complex can be extrapolated to estimate the effect of substituting a buried residue on protein stability.

To relate these thermodynamic data to detailed structural information, the crystal structures of the complexes between SGPB and the 20 coded P₁ variants of OMTKY3 are being determined and analyzed in our laboratory. I have chosen to determine the structures of 10 of the complexes, namely the Gly¹⁸, Ala¹⁸, Leu¹⁸, Glu¹⁸, Gln¹⁸, Asp¹⁸, Asn¹⁸, Pro¹⁸, Phe¹⁸ and Tyr¹⁸ variants in complex with SGPB.

Chapter 3 of this thesis reports the structures of the Gly¹⁸, Ala¹⁸ and Leu¹⁸ variants in complex with SGPB. These high resolution (Leu¹⁸ and Gly¹⁸ at 1.8Å, Ala¹⁸ at 1.4Å), very well refined structures (R-factors of 13.2%, 13.9% and 17.1% respectively) provide important insights for understanding the sequence-reactivity relationship of OMTKY3. Detailed comparisons among the three structures have revealed that water molecules redistribute themselves in adjusting to the size and the shape of the P₁ side-chain in the S₁ binding pocket of SGPB. Complexes of SGPB and four non-coded P₁ variants (Abu, Ape, Ahx and Ahp) have been modeled on the basis of the Ala¹⁸ structures. The homologous series of straight aliphatic side-chain (from 0 to 5 methylene or methyl groups) P₁ variants, Gly¹⁸, Ala¹⁸, Abu¹⁸, Ape¹⁸, Ahx¹⁸ and Ahp¹⁸, display a linear correlation between the free energy of binding and the hydrophobic surface area buried in the enzyme-inhibitor interface. The resulting correlation constant, 34 cal/mol/Å², has suggested that a modified value of hydrophobic strength, rather than that determined from

mutagenesis studies of protein cores ($45\text{--}60\text{ cal/mol/\AA}^2$), should be applied to the S_1 pocket, and probably to other protein-protein interfaces, because they are partially hydrated. By applying the correlation constant, the thermodynamic property of the complexes, K_a , can be related the structural property of the complex, the buried hydrophobic surface area, in a quantitative fashion.

Chapter 4 of this thesis describes the complexes of SGPB with the Glu¹⁸, Gln¹⁸ and Asp¹⁸ variants at pH 6.5 and 10.7, and Asn¹⁸ at pH 6.5. Experimental measurements of K_a have shown that the P₁ Glu18 and Asp18 display unusually high pK_a , 8.7 and 9.4 respectively, when bound in the very hydrophobic S_1 pocket of SGPB (Qasim et al., 1995). The structures of these complexes have been determined at the specified pHs, in an attempt to understand how the S_1 pocket copes with the negatively-charged carboxylate group. We unexpectedly found a K⁺ ion in the high pH complex of the Glu¹⁸ and Asp¹⁸ variants. It forms a salt bridge with one of the oxygens of the P₁ carboxyl group. The energy cost for accommodating this ion-pair in the hydrophobic pocket is about 5.6 Kcal/mol, as estimated from the difference in free energy of binding between the protonated and the negatively-charged Glu¹⁸ variant. The observed destabilizing effect of buried salt-bridges is contrary to the previously accepted view that buried salt-bridges stabilize proteins (Fersht, 1972; Anderson et al., 1990), and supports the recent results from a theoretical calculation (Hendsch & Tidor, 1994) and mutagenesis studies on the Arc repressor (Waldburger et al., 1995). Comparisons have also revealed the redistribution of water molecules to adjust to the subtle change of hydrophilicity of the P₁ side-chain in the S_1 pocket. The electrostatic effects on the free energy of binding of

these ionizable (Glu¹⁸, Asp¹⁸) and polar (Gln¹⁸, Asn¹⁸) P₁ variants are analyzed by the interactions the P₁ side-chains involved.

The crystal structures of the Pro¹⁸, Phe¹⁸ and Tyr¹⁸ variants have also been solved and refined at 1.8 Å resolution, although they will not be included in this thesis. As expected, the Pro¹⁸ variant is the worst inhibitor among all the P₁ variants, presumably due to the conformational peculiarity of this amino acid. However, as confirmed by the crystal structure, there is no frame-shift in the complex. The enzyme is still able to recognize the framework of the inhibitor despite the unfavorable P₁ residue. Therefore, it is an issue of molecular recognition and a test of the flexibility of the enzyme and the inhibitor. The crystallization of the complex, which is fairly weak ($K_d \sim 10^4 \text{ M}^{-1}$), was a challenge in itself.

Crystallographic aspects

X-ray crystallography is the most powerful technique yet available for visualizing macromolecular structures at the atomic level. The technique and its applications have been described in depth in several excellent books:

Glusker JP and Trueblood KN, *Crystal structure analysis. A primer*. New York: Oxford press, 1985; an introductory text for non-crystallographers.

Blundell LT and Johnson LN, *Protein crystallography*. Academic Press, 1976; a classic text on protein crystallography.

Stout GH and Jensen LH, *X-ray structure determination*. New York: Macmillan, 1968; on small molecular crystallography.

The spatial positions of protein atoms in a crystal are revealed by the X-ray diffraction of the crystal. The X-ray reflections are the the moduli of the Fourier transforms of the atomic arrangement in that crystal. The major task in crystallography is to determine the phase of each reflection. There are three conventional ways to determine the phases. Multiple Isomorphous Replacement (MIR) and Multiple Anomalous Dispersion (MAD) both involve the introduction or presence of heavy atoms to the protein of interest. The heavy atoms serve as references and the identification of the heavy atom binding sites leads to the identification of coordinates of protein atoms.

The Molecular Replacement method (MR) is the third common approach to determine phases (Rossmann, 1972). This is the technique by which the structures reported in this thesis were determined. The molecular replacement method utilizes the known structure of a protein, which is homologous to the protein of interest, to find out the orientation and position of the unknown structure in the crystal unit cell.

The Molecular replacement methods involve a rotation function search and a translation function search, in which the homologous protein of known structure is used as a search model. The rotation function search is used to find the orientation of the unknown structure in the unit cell, by doing an interatomic vector-pattern match. It is based on the assumption that the positions of protein atoms relative to one another (or the interatomic vectors) are similar in the two

proteins. The search model is rotated in the crystal unit cell until the maximum overlap of the two patterns is observed.

Knowing the orientation of the unknown structure, a translation search in the unit cell is carried out, in order to obtain the correct position for the unknown molecule. The optimal position is found when the crystallographic R-factor (an indication of the agreement between the observed and calculated structure factors) is at minimum and below an acceptable value (usually around 40%). It has to be further proven by an interpretable electron density map from which the entire polypeptide chain(s) can be traced unambiguously, and the resulting molecular model can be refined to a good R-factor and stereochemistry.

In practice, there are many technical modifications to help to distinguish the correct solution from the wrong ones. Quite often the molecular replacement method fails to give a correct solution, or does not permit a complete tracing of the polypeptide chain(s), mainly for the following reasons:

1. There is an insufficient degree of similarity between the model and the unknown protein. The interatomic vector pattern of the model is inconsistent with that of the unknown structure, so that a good match can not be possibly found or discerned above background.

2. There is significant inter-domain movement in the unknown structure relative to the search model, resulting in an interatomic vector pattern that is no longer homologous to that of the search model.

3. The search model is considerably incomplete relative to the unknown structure, where a significant portion of the molecular mass is

not represented by the search model. Hence large errors exist in the phase information derived solely from the search model.

In the determination of the structure the complex between elastase and the *Ascaris* inhibitor, native elastase was used as the search model for the molecular replacement. This model provided a correct solution, which resulted in an electron density map showing clear density for elastase molecules. However, only a small portion of the inhibitor (in the active-site region) could be identified, whereas the majority of the inhibitor could not be found in the electron density map. This is because the inhibitor, which constitutes about 20% of the molecular mass of the entire complex, was not accounted for by the search model. The phases determined solely from the elastase model were thus severely biased. The problem was overcome by using the electron density averaging technique (Vellieux et al., 1993), assisted by the fact that there are two copies of the complex in each asymmetric unit. By averaging their electron density, the signal to noise ratio of the density peaks was significantly elevated. Continuous density was revealed in previously featureless regions of the map, thereby allowing the unambiguous tracing of the entire inhibitor (except for the final three residues on the C-terminus).

In refining and interpreting the complexes of SGPB with the Glu¹⁸ and the Asp¹⁸ variants, one of the limitations of protein crystallography, the resolution limit, became a big concern. The charge state of the Glu¹⁸ and Asp¹⁸ could not be determined from looking at the electron map, since one can not differentiate a -OH group from an oxygen atom, which differs by only one proton, at 1.8 Å resolution. The

difference in bond length between the C=O and C-O bonds, which is only 0.1 Å, is comparable to the coordinate errors of a well-refined 1.8 Å resolution structure. Thus it is difficult to distinguish if the electron density peak is associated with a -OH group or an oxygen atom. The final decision has to be assisted by other available information, such as the interactions with other atoms.

REFERENCES

- Anderson DE, Becktel WJ & Dahlquist FW. 1990. pH-induced denaturation of proteins: a single salt-bridge contributes 3-5 kcal/mol to the free energy of folding of T4 lysozyme. *Biochemistry* 29:2403-2408.
- Ardelt WA & Laskowski M Jr. 1983. Thermodynamics and kinetics of the hydrolysis and resynthesis of the reactive site peptide bond in turkey ovomucoid third domain by aspergillopeptidase B. *Acta Biochem Polon* 30:115-126.
- Barrett AJ. 1986. An introduction to the proteinases. In *Proteinase inhibitors*. Barrett AJ and Salvesen G, eds, Amsterdam: Elsevier. pp 1-22.
- Berger A & Schetcher I. 1970. Mapping the active site of papain with the aid of peptide substrates and inhibitors. *Phil Trans R Soc London* B257:249-264.
- Babin, D. R., Peanasky, R. J. & Goss, S. M. (1984). The isoinhibitors of chymotrypsin/elastase from *Ascaris lumbricoides*: the primary structure. *Arch. Biochem. Biophys.* 232, 143--161.
- Bigler TL, Lu W, Park SJ, Tashiro M, Wieczorek WR & Laskowski M Jr. 1993. Binding of amino acid side chains to preformed cavities: Interaction of serine proteinases with turkey ovomucoid third domains with coded and noncoded P₁ residues. *Protein Science* 2:786-799.
- Blanchard H & James MNG. 1995. X-ray crystallographic structure of *Streptomyces griseus* proteinase B at its enzymatically optimal pH. Manuscript in preparation.
- Bode W, Chen Z, Bartels K, Kutzbach C, Schmidt-Kastner G & Bartunik H. 1983. Refined 2 Å X-ray crystal structure of porcine pancreatic kallikrein A, a specific trypsin-like serine proteinase -

crystallization, structure determination, crystallographic refinement, structure and its comparison with bovine trypsin. *J Mol Biol* 164:237-282.

Bode W, Epp O, Huber R, Laskowski M Jr. & Ardelt W. 1985. The crystal and molecular structure of the third domain of silver pheasant ovomucoid (OMSVP3). *Eur J Biochem* 147:387-395.

Bode W, Wei AZ, Huber R, Meyer E, Travis J & Neumann S. 1986. X-ray crystal structure of the complex of human leukocyte elastase (PMN elastase) and the third domain of the turkey ovomucoid inhibitor. *EMBO J* 5:2453-2458.

Bode W, Mayr I, Baumann U, Huber R, Stone SR & Hofsteenge J. 1989a. The refined 1.9 Å crystal structure of human α -thrombin: Interaction with D-Phe-Pro-Arg chloromethyl ketone and significance of the Tyr-Pro-Pro-Trp insertion segment. *EMBO J* 8:3467-3475.

Bode W, Meyer E Jr & Powers JC. 1989b. Human leukocyte and porcine pancreatic elastase: X-ray crystal structures, mechanism, substrate specificity, and mechanism-based inhibitors. *Biochemistry* 28:1951-1963.

Bode W & Huber R. 1992. Natural protein proteinase inhibitors and their interactions with proteinases. *Eur J Biochem* 204:433-451.

Brayer GD, Delbaere LTJ & James MNG. 1979. Molecular structure of the α -lytic protease from *Mycobacter* 495 at 2.8 Å resolution. *J Mol Biol* 131:743-775.

Brown FJ, Andisik DW, Bernstein R, Bryant CB, Ceccarelli C, Damewood JR Jr, Edwards PD, Earley RA, Reeney S, Grfeen RC et al. Design of orally active, non-peptidic inhibitors of human leukocyte elastase. *J Med Chem* 37:1259-1261.

Carter P & Wells JA. 1988. Dissecting the catalytic triad of a serine protease. *Nature* 332:564-568.

- Codding PW, Delaere LT, Hayakawa K, Hutcheon WL, James MNG & Jurasek L. 1974. The 4.5 Angstrom resolution structure of a bacterial serine proteinase from *Streptomyces griseus*. *Can J Biochem* 52:208.
- Correy DR & Craik CS. 1992. An investigation into the minimum requirements for peptide hydrolysis by mutation of the catalytic triad of trypsin. *J Am Chem Soc* 114:1784-1790.
- Craik CS, Roxzniak S, Largman C & Rutter WJ. 1987. The catalytic role of the active site aspartic acid in serine protease. *Science* 237:909-913.
- Delbaere LTJ, Brayer GD & James MNG. 1979. The 2.8 Å resolution structure of *Streptomyces griseus* protease B and its homology with α-chymotrypsin and *Streptomyces griseus* protease A. *Can J Biochem* 57:135-144.
- Dixon MM, Brennan RG & Matthews BW. 1991. Structure of γ-chymotrypsin in the range pH 2.0 to pH 10.5 suggests that γ-chymotrypsin a covalent acyl-enzyme adduct at low pH. *Int J Biol Macromol* 13:89-96.
- Ferth AR. 1972. Conformational equilibria in α- and δ- chymotrypsin . The energetics and importance of the salt bridge. *J Mol Biol* 64:497-509.
- Fujinaga M, Delbaere LTJ, Brayer GD & James MNG. 1985. Refined structure of α-lytic protease at 1.7 Å resolution. Analysis of hydrogen bonding and solvent structure. *J Mol Biol* 183:479-502.
- Fujinaga M & James MNG. 1987a. Rat submaxillary gland serine protease, tonin - structure solution and refinement at 1.8 Å resolution. *J Mol Biol* 195:373-397.
- Fujinaga M, Sielecki AR, Read RJ, Ardelt W, Laskowski M Jr. & James MNG. 1987b. Crystal and molecular structures of the complex of α-chymotrypsin with its inhibitor turkey ovomucoid third domain at 1.8 Å resolution. *J Mol Biol* 195:397-418.
- Greenblatt, H. M., Ryan, C. A. & James, M. N. G. (1989). Structure of the complex of *Streptomyces griseus* proteinase B and polypeptide

chymotrypsin inhibitor-1 from Russet Burbank potato tubers at 2.1Å.
J. Mol. Biol. 205, 201--228.

Goodman, R. B. & Peanasky, R. J. (1982). Isolation of the trypsin inhibitors in *Ascaris Lumbricoides* var suum using affinity chromatography. *Anal. Biochem.* 120, 387--393.

Harel M, Su CT, Frolow F, Silan I & Sussman JL. 1991. g-chymotrypsin is a complex of a-chymotrypsin with its own autolysis products. *Biochemistry* 30:5217-5225.

Hartley BS. 1960. Proteolytic enzymes. *Annu Rev Biochem* 29:45-72.

Hartley BS. 1970. Homologies in serine proteinases. *Phil Trans R Soc Lond B* 257:77-87.

Harper JW, Cook RR, Roberts CJ, MaLaughlin BJ, Powers JC. 1984. Active site mapping of the serine proteases human leukocyte elastase, cathepsin G, porcine pancreatic elastase, rat mast cell proteases I and II, bovine chymotrypsin α and *S.aureus* protease V-8 using tripeptide thiobenzyl ester substrates. *Biochemistry* 23:2995-3002.

Hedstrom L, Szilagyi L, Rutter WJ. 1992. Converting trypsin to chymotrypsin: The role of surface loops. *Science* 255:1249-1253.

Hedstrom L, Farr-Jones S, Kettner CA & Rutter WJ. 1994. Converting trypsin to chymotrypsin: Ground state binding does not determine substrate specificity. *Biochem* 33:8764-8769.

Hendsch ZS & Tidor B. 1994. Do salt bridges stabilize proteins? a continuum electrostatic analysis. *Protein Sci* 3:211-226.

James MNG. 1976. Relationship between the structures and activities of some microbial serine proteases. II. Comparison of the tertiary structure of microbial and pancreatic serine proteases. In: Ribbons DW and Brew K, eds. *Proteolysis and Physiological Regulations*. New York: Academic Press. pp 125-142.

- James MNG, Sielecki AR, Brayer GD, Delbaere LTJ & Bauer CA. 1980. Structures of product and inhibitor complexes of *Streptomyces griseus* protease A at 1.8 Å: A model for serine protease catalysis. *J Mol Biol* 144:43-49.
- James MNG. 1994. Convergence of active-centre geometries among the proteolytic enzymes. Bond JS & Barrett AJ, eds. In *Proteolysis and protein turnover*. Brookfield, VT: Portland press. pp 1-8.
- Jurasek L, Johnson P, Olafson RW & Smillie LB. 1971. An improved fractionation system for pronase on CM-sephadex. *Can J Biochem* 49:1195-1201.
- Jurasek L, Carpenter MR, Smillie LB, Gertler A, Levy S & Ericsson, LH. 1974. Amino acid sequence of *Streptomyces griseus* protease B, a major component of pronase. *Biochem Biophys Res Commun* 61:1095-1100.
- Komiyama T, Bigler TL, Yoshida N, Noda K & Laskowski M Jr. 1991. Replacement of P₁ Leu¹⁸ by Glu¹⁸ in the reactive-site loop of turkey ovomucoid third domain converts it into a strong inhibitor of Glu-specific *Streptomyces griseus* proteinase (GluSGP). *J Biol Chem* 266:10727-10730.
- Laskowski M Jr. & Kato I. 1980. Protein inhibitors of proteinases. *Biochemistry* 49:593-626.
- Liao D, Breddam K, Sweet RM, Bullock T & Remington SJ. 1992. Refined atomic model of wheat serine carboxypeptidase II at 2.2 Å resolution. *Biochemistry* 31:9796-9812.
- Lu W, Zhang W, Molly SS, Thomas G, Ryan K, Chiang Y, Anderson S & Laskowski M Jr. 1993. Arg¹⁵-Lys¹⁷-Arg¹⁸ turkey ovomucoid third domain inhibits human furin. *J Biol Chem* 268:14583-14585.
- Lu W. 1994. Energetics of the interactions of ovomucoid third domain variants with different serine proteinases. *Ph.D. Thesis*, Purdue University, West Lafayette, Indiana.

- Matthews BW, Sigler PB, Henderson R, Blow DM. 1967. Three-dimensional structure of tosyl- α -chymotrypsin. *Nature* 214:652-656.
- Mattos C, Rasmussen B, Ding X, Petsko GA & Ringe D. 1994. Analogous inhibitors of elastase do not always bind analogously. *Nature Struct Biol* 1:55-58.
- Martzen, M. R., Geise, G. L., Hogan, B. J. & Peanasky, R. J. (1985). *Ascaris suum*: Localization by immunochemical and fluorescent probes of host proteases and parasite proteinase inhibitors in cross-sections. *Expl. Parasitol.* 60, 139--149.
- Narajana SV, Carson M, el-Kabbiani O, Kilpatrick JM, Moore D, Chen X, Bugg CE & DeLucs LJ. 1994. Structure of human factor D. A complement system protein at 2.0 Å resolution. *J Mol Biol* 235:695-708.
- Navia MA, McKeever Bm, Springer JP, Lin TY, Williams HR, Fluder EM, Dorn CP & Hoogsteen K. 1989. Structure of human neutrophil elastase in complex with a peptide chloromethyl ketone inhibitor at 1.84Å resolution. *Proc Natl Acad Sci USA* 86:7-11.
- Nienaber VL, Breddam K & Birktoft JJ. 1993. A glutamic acid specific serine protease utilizes a novel histidine triad in substrate binding. *Biochemistry* 32:11469-11475.
- Padmanabhan K, Padmanabhan KP, Tulinsky A, Park CH, Bode W, Huber R, Blankenship DR, Cardin AD, Kisiel W. 1993. Structure of human ds(1-45) factor Xa at 2.2 Å resolution. *J Mol Biol* 232:947-966.
- Papamokos E, Weber E, Bode W, Empie MW, Kato I & Laskowski M Jr. 1982. Crystallographic refinement of Japanese quail ovomucoid, a-Kazal type inhibitor, and model building studies of complexes with serine proteases. *J Mol Biol* 158:515- 537.
- Peanasky, R. J., Bentz, Y., Paulson, B., Graham, D. L. & Babin, D. R. (1984a). The isoinhibitors of chymotrypsin/elastase from *Ascaris Lumbricoides*: Isolation by affinity chromatography and association with the enzymes. *Arch. Biochem. Biophys.* 232, 127--134.

- Peanasky, R. J., Bentz, Y., Homandberg, G. A., Minor, S. T. & Babin, D. R. (1984b). The inhibitors of chymotrypsin/elastase from *Ascaris Lumbricoides*: The reactive site. *Arch. Biochem. Biophys.* 232, 135-142.
- Peanasky, R. J., Martzen, M. R., Homandberg, G. A., Cash, J. M., Babin, D. R. & Litweiler, B. (1987). Proteinase inhibitors from intestinal parasitic helminths: Structure and indications of some possible functions. In *Molecular paradigms for eradicating helminthic parasites* (MacInnis, A. J., ed.), pp. 349--366, DPL, New York.
- Perona JJ, Craik CS & Fletterick Rj. 1993. Locating the catalytic water molecule in serine proteinases. *Science* 261:620-621.
- Perona JJ & Craik CS. 1995. Structural basis of substrate specificity in the serine proteases. *Protein Sci* 4:337-360.
- Poulos TL, Alden RA, Freer ST, Birktoft JJ & Kraut J. 1976. *J Biol Chem* 251:1097-1103.
- Qasim MA, Ranjbar MR, Wynn R, Anderson S & Laskowski M Jr. 1995. Ionizable P₁ residues in serine proteinase inhibitors undergo large pK_a shifts on complex formation. *J Biol Chem* (in press).
- Read RJ, Fujinaga M, Sielecki AR & James MNG. 1983. Structure of the complex of *Streptomyces griseus* protease B and the third domain of the turkey ovomucoid inhibitor at 1.8 Å resolution. *Biochemistry* 22:4420-4433.
- Read RJ & James MNG. 1986. Introduction to the protein inhibitors: X-ray crystallography. In *Proteinase inhibitors* (Barrett AJ and Salvesen G, Eds.) pp. 301-336. Elsevier, Amsterdam.
- Remington SJ, Woodbury RG, Reynolds RA, Matthews BW & Neurath H. 1988. The structure of rat mast cell protease at 1.9 Å resolution. *Biochemistry* 27:8097-8105.
- Rossmann MG. 1972, *The molecular replacement method*. International science review 13, Gordon & Breach, New York.

- Singer PT, Smalas A, Carty RP, Mangel WF & Sweet RM. 1993. The hydrolytic water molecule in trypsin, revealed by time-resolved Laue crystallography. *Science* 259:669-673.
- Schechter I & Berger A. 1967. On the size of the active site in proteases. I. Papain. *Biochem Biophys Res Commun* 27:157-162.
- Tsu CA, Perona JJ, Schelllenberger V, Turck CW & Craik CS. 1994. The substrate specificity of *Ucs pugilator* collagenolytic serine protease 1 correlates with the bovine type I collagen cleavage sites. *J Biol Chem* 269:19565-19572.
- Vellieux FMD, Hajdu J, Verlinde LMJ, Groendijk H, Read RJ, Greenhough TJ, Campbell JW, Kalk KH, Littlechild JA, Watson HC & Hol WGJ. 1993. Structure of glycosomal glyceraldehyde-3-phosphate dehydrogenase from *Trypanosoma brucei* determined from Laue data. *Proc Natl Acad Sci USA* 90:2355-2359.
- Waldburger CD, Schildbach JF & Sauer RT. 1995. Are buried salt bridges important for protein stability and conformational specificity?. *Nature Struct Biol* 2:122-128.
- Watson HC, Shotton DM, Cox JC, Muirhead H. 1970. Three-dimensional fourier synthesis of tosyl elastase at 3.5 Å resolution. *Nature* 225:806-811.
- Weber E, Papamokos E, Bode W, Huber R, Kato I. & Laskowski M Jr. 1981. Crystallization, crystal structure analysis and molecular model of the third domain of Japanese Quail ovomucoid, a Kazal type inhibitor. *J Mol Biol* 149:109-123.
- Wieczorek M & Laskowski M Jr. 1983. Formation of covalent hybrids from amino- terminal and carboxy-terminal fragments of two ovomucoid third domains. *Biochemistry* 22:2630-2636.
- Wieczorek M, Park SJ & Laskowski M Jr. 1987. Covalent hybrids of ovomucoid third domains made from one synthetic and one natural peptide chain. *Biochem Biophys Res Commun* 144:499-504.

Wright CS, Alden RA & Kraut J. 1969. Structure of subtilisin BPN' at 2.5 Å resolution. *Nature* 221:235-242.

Zhou GW, Guo J, Huang W, Fletterick RJ, Scanlan TS. 1994. The three dimensional structure of a catalytic antibody with active site similarity to serine protease. *Science* 265:1059-1064.

CHAPTER 2

The Molecular Structure of the Complex of the
Chymotrypsin/Elastase Inhibitor from
Ascaris Lumbricoides Suum with Porcine Elastase*

INTRODUCTION

Ascaris is a parasitic nematode that resides within the intestines of infected humans (*Ascaris lumbricoides*) and hogs (*Ascaris suum*). The two species are morphologically indistinguishable but show species specificity. The presence of this parasite in children can be an important cause of malnutrition. *Ascaris* produces small proteins, between 60 and 150 residues, which inactivate chymotrypsin and elastase (Peanasky et al., 1984a), trypsin (Goodman & Peanasky, 1982), pepsin (Abuereish & Peanasky, 1974) and carboxypeptidase (Homandberg & Peanasky, 1976). The physiological roles of these inhibitory proteins is to protect the worms from proteolytic degradation by the host's digestive enzymes. Studies have shown that the worms do not secrete the inhibitors; instead, they take up the host's proteinases through their intestine (Mart et al., 1985). The host's proteinases are present inside the worm as inactive enzyme-inhibitor complexes. *Ascaris* cannot inhibit plant cysteine proteinases such as papain and

*A version of this chapter has been published. Huang et al., 1994. Structure 2: 679-689.

ficin. Incubation of live worms with papain destroys their intestinal tract and then the entire worm (Peanasky et al., 1987).

Eight different proteins from *Ascaris lumbricoides* suum have been identified as inhibitors of serine proteinases; there are five chymotrypsin or elastase inhibitors and three trypsin inhibitors. These inhibitors have between 62 and 65 residues, and share 30 to 40% sequence identity with each other (Babin et al., 1984). There are five disulfide bonds present in each inhibitor. These inhibitors can be grouped into a novel family of *Ascaris* inhibitors, as they display no sequence identity with the other ten or more inhibitor families of serine proteinase (reviewed by Laskowski & Kato, 1980). Much is known about the protein inhibitors from other families. Detailed thermodynamic association constants for ovomucoid inhibitor variants with several serine proteinases have been determined (Bigler et al., 1993). Crystal structures have been reported for bovine pancreatic trypsin inhibitor (Kunitz-Kunin family; Marquart et al., 1983; Wlodawer et al., 1987), avian ovomucoid inhibitor third domain (Kazal family; Read et al., 1983; Fujinaga et al., 1987; Bode et al., 1986; Papamokos et al., 1982; Bode et al., 1985), *Streptomyces* subtilisin inhibitor (SSI family; Hirono et al., 1984), soybean trypsin inhibitor and *Erythrina* trypsin inhibitor (STI-Kunitz family; Sweet et al., 1974; Onesti et al., 1991), barley seed chymotrypsin inhibitor 2 (McPhalen et al., 1985a; McPhalen & James, 1987) and leech inhibitor eglin-C (PI-1 family; McPhalen et al., 1985b; McPhalen & James, 1988; Bolognesi et al., 1990; Dauter & Bet, 1991), polypeptide chymotrypsin inhibitor from potato (PI-2 family; Greenblatt et al., 1989), acid-

stable inhibitor from human mucous secretions (Chelonianin family; Grutter et al., 1988), Bowman-Birk inhibitors (Chen et al., 1992; Tshuogae et al., 1986; Lin et al., 1993), squash seed trypsin inhibitor (squash seed inhibitor family; Bode et al., 1989a), serpins (Lobermann et al., 1984; Delarue et al., 1990; Baumann et al., 1991; Wright et al., 1990; Stein et al., 1990; Schreuder et al., 1994; Carrell et al., 1994) and hirudin (Rydel et al., 1990; Grutter et al., 1990). The inhibitors from these different families have a common conformation for the reactive-site loop, while displaying completely different overall structures (reviewed by Read & James, 1986; Bode & Huber, 1992).

An NMR study has reported the secondary structural features of one of the trypsin inhibitors from *Ascaris* (ATI) (Gronenborn et al., 1990). In it, two regions of β -sheet were identified. In order to provide a complete tertiary structure and to establish conclusively the disulfide bridge connections, we have undertaken the crystal structure determination of one of the chymotrypsin/elastase inhibitors, designated as the C/E-1 inhibitor. The C/E-1 inhibitor has 63 amino acids and shares 40% sequence identity with ATI. It forms very tight complexes with chymotrypsin and pancreatic elastase, with dissociation constants of 3.8×10^{-12} M and 6.3×10^{-11} M, respectively (Peanasky et al., 1984a). It has been demonstrated that the reactive-site peptide bond (the scissile bond) is between Leu31 and Met32 (Peanasky et al., 1984b).

The C/E-1 inhibitor has been cocrystallized with porcine pancreatic elastase (PPE). Elastase is an important enzyme in several regards (Bode et al., 1989b). Pancreatic elastase is of medical relevance, being implicated in pancreatitis. The closely related human leukocyte elastase (40% sequence identity with PPE) has been associated with a number of inflammatory disorders including pulmonary emphysema, acute respiratory distress syndrome, rheumatoid arthritis and cystic fibrosis. The crystal structure of PPE was among the first serine proteinase structures to be determined (Shotton & Watson, 1970). The difficulty we experienced in crystallizing the C/E-1 inhibitor alone was overcome by co-crystallizing it with PPE. In addition, the molecular replacement technique using native elastase (Meyer et al., 1988) as the search model, greatly facilitated the structure determination of the C/E-1 inhibitor and obviated the need for any heavy atom derivatives.

RESULTS AND DISCUSSION

Overall Structure of the C/E-1 Inhibitor

There are two copies of the enzyme-inhibitor complex, displaying nearly identical conformations, in the asymmetric unit of the crystals studied. Pair-wise superpositions of the C α atoms of the two elastase molecules (240 atoms) and the two inhibitor molecules (61 out of 63 atoms) give root mean square (rms) differences of 0.35 Å and 0.38 Å, respectively. An rms difference of 0.4 Å for all C α atoms is obtained

by superimposing the two copies of the entire complex, indicating that the binding modes of the two inhibitors are virtually indistinguishable. Therefore, the discussion of the structure will not discriminate between each complex unless otherwise stated.

The C/E-1 inhibitor folds into a wedge-shaped disc as illustrated in Figures 2.1 and 2.2a. The diameter of the disc is approximately 30 Å and the thickness is about 15Å on the thicker edge. The secondary-structural elements and the covalent pairings of the five disulfide bridges are shown schematically in Figure 2.2b. In these figures and resulting discussion the inhibitor residues are labeled with an "I" following the sequence number to distinguish them from the residues of elastase.

The secondary structure assignment is based on the criteria of Kabsch and Sander (Kabsch & Sander, 1983) as well as on the main-chain hydrogen-bonding patterns. The criteria used to define a hydrogen-bonding interaction were: the maximum donor to acceptor distance was 3.4Å and the maximum hydrogen to acceptor distance was 2.4Å; the angle defined by the donor atom-hydrogen atom-acceptor atom should be greater than 135°. The major secondary structures of the C/E-1 inhibitor are two β -sheets. β -sheet 1 comprises three strands, strand 2 (Met19I to Cys21I), strand 4 (Gly45I to Thr49I) and strand 5 (Lys53I to Ala57I). In fact, strands 4 and 5 form a twisted β -hairpin, while strand 2 runs parallel to strand 4. Only two hydrogen-bonds are formed between the β -hairpin and the otherwise isolated strand 2. The hydrogen-bonding pattern for this β -sheet is shown in Figure 2.3. β -

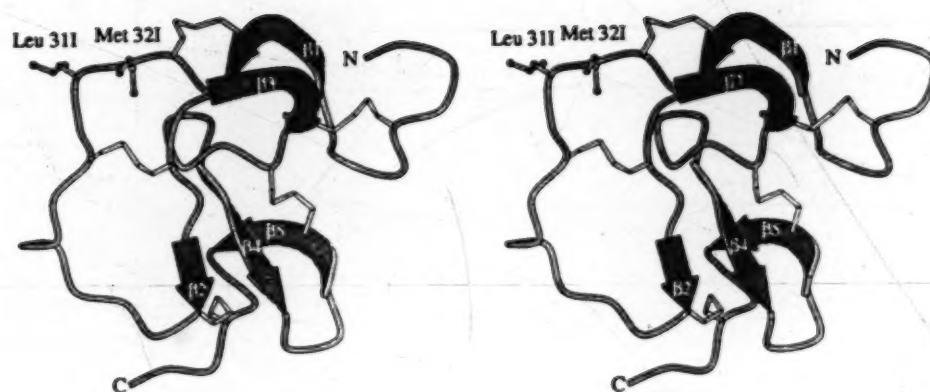


Figure 2.1 Schematic representation of the C/E-1 inhibitor. Arrows represent β -strands. Disulfide bridges and the side-chains of the P_1 and P_1' residues Leu31I and Met32I are shown. The peptide bond Leu31I-Met32I is the scissile bond. (Diagram generated using MOLSCRIPT, Kraulis, 1991).

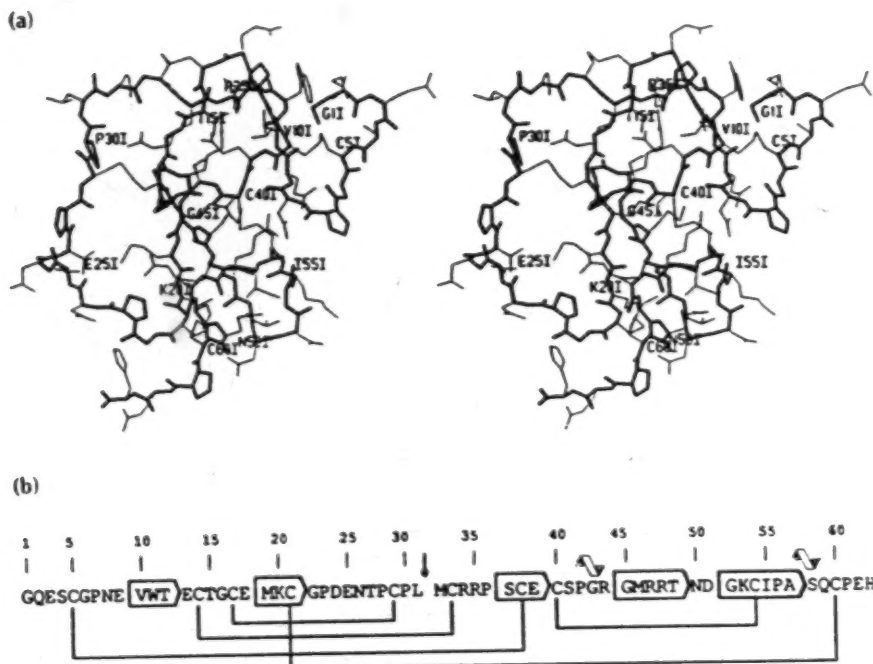


Figure 2.2 (a). An all-atom representation of the C/E-1 inhibitor, from approximately the same view as Figure 2.1. Main-chain atoms are shown in thick lines, side-chain atoms are shown in thin lines. Every fifth residue is labeled. The carboxy-terminal residues, Glu62I and His63I have poor electron density and are shown here in an arbitrary conformation.

(b). Sequence, secondary structure and disulfide bridges of the C/E-1 inhibitor. Residues in β -strands are enclosed in arrow-shaped boxes. The two single turn 3_{10} -helices (Pro42I to Arg44I and Ala57I to Gln59I) are denoted by the helical symbol above the sequence letters of the residue involved. The vertical arrow indicates the scissile peptide bond.

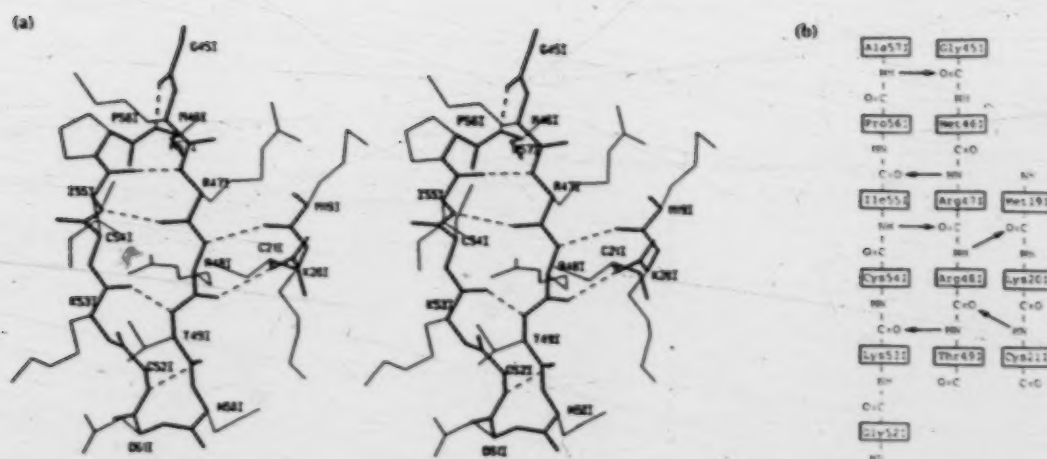


Figure 2.3 (a). Hydrogen-bonding patterns in β -sheet 1. Main-chain atoms are shown in thick lines, side-chain residues are shown in thin lines. Dashed lines represent possible hydrogen bonds.

(b). Schematic representation of the inter-strand hydrogen bonds of β -sheet 1

sheet 2 consists of strand 1 (Val10I to Thr12I) and strand 3 (Ser37I to Glu39I). In total, four hydrogen bonds are formed between the two antiparallel β -strands. Immediately adjacent to β -sheet 2, there is a single-residue β -ladder (Thr15I and Arg34I), having two hydrogen-bonds formed between these two residues. Details of the hydrogen-bonding interactions in β -sheet 2 and the small β -ladder are shown in Figure 2.4.

In contrast to most inhibitors of known structure, there are no extensive helical regions in the C/E-1 inhibitor. Instead, only two single-turn 3_{10} -helices can be identified. One such 3_{10} -helix (Ala57I to Gln59I) is located near the C-terminus of the C/E-1 inhibitor. The other (Pro42I to Arg44I) is between β -strands 3 and 4. Interestingly, β -structures are more commonly observed than α -helices among protein inhibitors of serine proteinases. While there are no all- α structures, there are all- β structures, such as *Erythrina* trypsin inhibitor (Oneesti et al., 1991), Mung bean trypsin inhibitor (Lin et al., 1993), mucous proteinase inhibitor (SLPI; Grutter et al., 1988) and squash seed trypsin inhibitor-1 (CMTI-I; Bode et al., 1989a).

The crystallographically-determined hydrogen-bonding patterns in the C/E-1 inhibitor agree in general with those of ATI as observed by the NMR studies (Cronenborn et al., 1990). The sequence alignment of the C/E-1 inhibitor with ATI is shown in Figure 2.5. The two hydrogen-bonds defining the two 3_{10} -helices are not reported for ATI by the NMR studies, nor are the two hydrogen bonds defining the small β -ladder between Thr15I and Arg34I in the C/E-1 inhibitor (equivalent

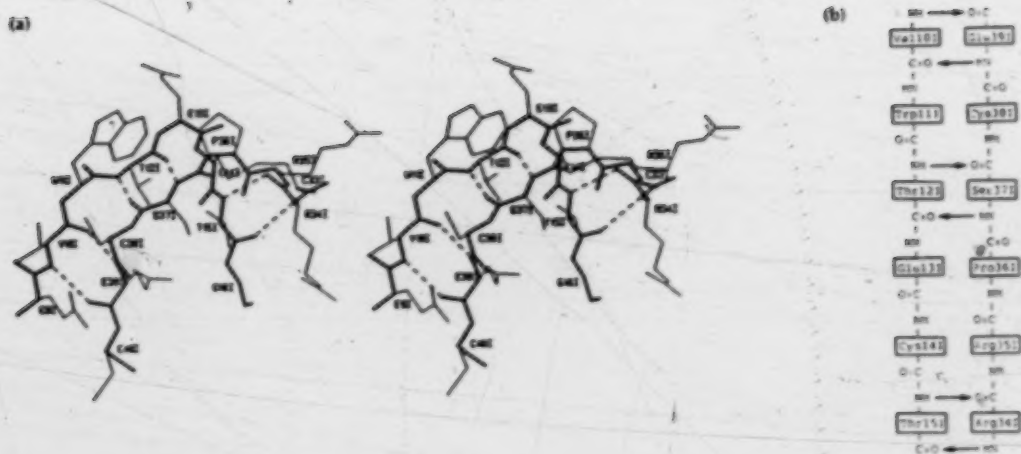


Figure 2.4 (a). Hydrogen-bonding patterns in β -sheet 2 and the small β -ladder. Main-chain atoms are shown in thick lines, side-chain residues are shown in thin lines. Dashed lines represent possible hydrogen bonds.

(b). Schematic representation of inter-strand hydrogen bonds in β -sheet 2 and the β -ladder.

C/E-1	1	G	Q	E	S	C	G	-	R	N	E	V	W	T	E	C	T	G	C	E	M	K	C	G	P	D	E	N	T	P	C	P	L	30
C/E-2		-	R	K	P	C	G	-	K	N	E	V	W	T	E	C	T	G	C	E	M	L	K	C	G	N	D	E	N	T	P	C	A	L
C/E-3		-	R	K	P	C	G	-	K	N	E	V	W	T	E	C	T	G	C	E	M	L	K	C	G	N	D	E	N	T	P	C	A	L
C/E-4		-	R	K	P	C	G	-	K	N	E	V	W	T	E	C	T	G	C	E	M	L	K	C	G	N	D	E	N	T	P	C	A	L
C/E-5		-	R	K	P	C	G	-	K	N	E	V	W	T	E	C	T	G	C	E	M	L	K	C	G	N	D	E	N	T	P	C	A	L
ATI		E	A	E	K	C	T	K	P	N	E	Q	W	T	K	C	G	G	C	E	M	G	T	C	A	Q	-	K	I	V	P	C	T	R
C/E-1	35	M	C	R	P	P	S	C	E	C	S	P	G	R	G	M	R	R	T	N	D	G	K	C	I	P	A	S	Q	C	P	E	H	60
C/E-2		M	C	R	P	P	S	C	E	C	S	P	G	R	G	M	R	R	T	N	D	G	K	C	V	P	V	S	E	C	P	R	K	M
C/E-3		M	C	R	P	P	S	C	E	C	S	P	G	R	G	M	R	R	T	N	D	G	K	C	V	P	V	S	E	C	P	R	K	M
C/E-4		M	C	R	P	P	S	C	E	C	S	P	G	R	G	M	R	R	T	N	D	G	K	C	V	P	V	S	E	C	P	R	K	M
C/E-5		M	C	R	P	P	S	C	E	C	S	P	G	R	G	M	R	R	T	N	D	G	K	C	V	P	V	S	E	C	P	R	K	M
ATI		E	C	K	P	P	R	C	E	C	I	A	S	A	G	F	V	R	D	A	Q	G	N	C	I	K	F	E	D	C	P	X		

Figure 2.5 Sequence alignment of the five homologous *Ascaris* C/E inhibitors with ATI. Conserved residues are in bold.

residues Gly15I and Lys34I in ATI). In addition, the peptide oxygen of Asp49I and the peptide nitrogen of Asn53I form a hydrogen bond in ATI, but the corresponding atoms from Thr49I and Lys53I diverge apart in the C/E-1 inhibitor. Nevertheless, it seems that the molecular scaffold of the C/E-1 inhibitor is already established in its free form in solution, and is largely retained in the crystal lattice of the enzyme-inhibitor complex.

The other major secondary structural features present in the C/E-1 inhibitor are reverse turns. Three turns are present including two type I turns and one type II turn. A type I turn (Glu2I to Cys5I) and a type II turn (Glu6I to Glu9I) are located at the N-terminus; the other type I turn (T49I to G52I) connects β -strands 4 and 5.

Solvent accessible surface calculations reveal that there is no clustering of hydrophobic atoms to form a fully buried hydrophobic core in the C/E-1 inhibitor. The absence of a hydrophobic core is understandable given the fact that only 25% of the residues of the C/E-1 inhibitor are hydrophobic residues, and half of these are proline. The relatively large number of disulfide bridges thus serves as an alternative to a hydrophobic core for maintaining a well-defined tertiary structure. The pairing of the 10 cysteine residues is shown in Figure 2.2b. Interestingly, except for the disulfide bridge formed by Cys17I and Cys29I, which is proximal to the N-terminal side of the reactive-site peptide bond (L31I-M32I), the other four disulfide bridges are all involved in cross-linking potentially flexible loops to the relatively well-defined β -sheet scaffold (Figure 2.1). For

example, Cys38I from strand 3 forms a disulfide bridge with Cys5I from a type I turn near the N-terminus, thereby imposing constraints on the conformation of the N-terminus. Similarly, the single-turn 3_{10} helix near the C-terminus is clamped to strand 2 by the disulfide bridge Cys21I to Cys60I. The stabilizing effects of this disulfide bridge on the conformation of the C-terminus can be appreciated by the fact that Pro61I, the residue immediately following Cys60I, has only weak electron density and the two carboxyl-terminal residues, Glu62I and His63I, have no recognizable electron density associated with them. The disulfide bond Cys40I-Cys54I attaches the loop between Glu39I and Gly45I to strand 5. The disulfide bridge connecting Cys14I to Cys33I brings Thr15I and Arg34I of the small β -ladder close in space, so as to resume the inter-strand hydrogen-bonding interactions interrupted by Pro36I following β -strand 3.

Such disulfide bridge-enhanced structures are also observed in the squash seed trypsin inhibitor (Bode et al., 1989a) and the mucous proteinase inhibitor (Clare et al., 1991). In C/E-1, CMTI-I and SLPI the frequency of occurrence of cysteine residues is approximately equal. However, the disposition of the disulfide bridges are completely different in these three inhibitor families, as are the overall tertiary structures.

In addition to the interconnections between secondary structural elements introduced by the disulfide bridges, electrostatic and hydrogen-bonding interactions involving side-chains also contribute substantially to maintain the integrity of the overall structure. The

hydrogen-bonding potentials of side-chains are well satisfied for most residues in the C/E-1 inhibitor. Prominent interactions are made by Arg48I (Figure 2.6) and Arg44I (Figure 2.7). Projecting out of strand 4, the side-chain of Arg48I forms hydrogen-bonds and/or salt-bridges with the peptide oxygen of Gly52I and the carboxylate oxygen atoms of Glu9I and Glu18I. These glutamate residues also form hydrogen bonds to the peptide nitrogen atoms of Gly6I and Cys40I, respectively. Through this electrostatic/hydrogen-bond network, connections are established among the β -sheets and the three loops, i.e. the loop preceding strand 1, the loop between strand 1 and strand 2 and the loop linking strand 3 and strand 4. On the obverse face of the disc, Arg44I links a reverse turn (Ser41I to Arg44I) to strand 3 by salt bridges or hydrogen-bonds to Glu39I and Ser41I.

Geometry of the Reactive Site Loop

A close-up view of part of the reactive-site loop of the C/E-1 inhibitor and the active site region of elastase is shown in Figure 2.8, superimposed onto the electron density map computed with coefficients derived from SIGMAA (Read, 1986). As expected, the reactive-site loop (residue Asn26I to Cys33I) of the C/E-1 inhibitor exhibits an extended conformation, closely resembling those of other protein inhibitors. Figure 2.9 shows the superposition of the reactive-site loops from several protein inhibitors of serine proteinases. The rms differences in main-chain atoms from residues P₃ to P₂' (nomenclature of Schechter & Berger, 1967) are summarized in Table 2.1.

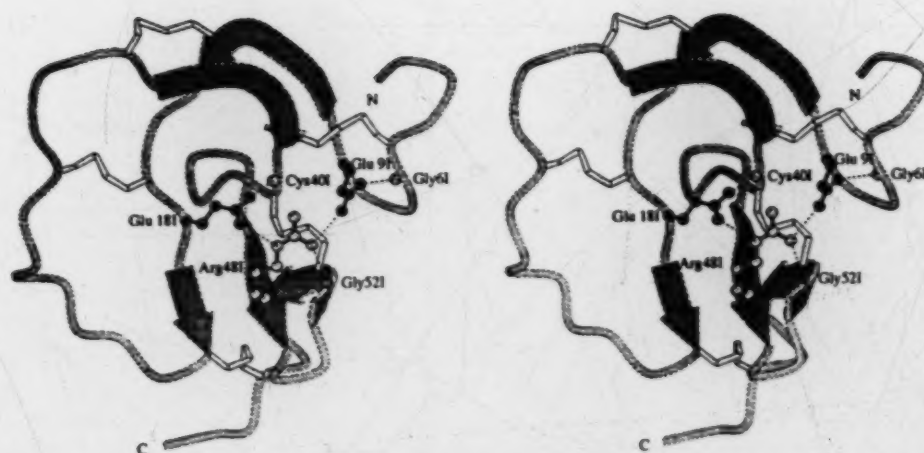


Figure 2.6 The salt bridge/hydrogen-bond network centered on Arg48I. The side-chain of Arg48I, and the peptide nitrogen atoms of Gly6I and Cys40I are shown in light grey. The side-chains of Glu9I and Glu18I and the peptide oxygen of Gly52I are shown in dark grey. Dashed lines represent hydrogen bonds or salt bridges within 3.2 Å.



Figure 2.7 The salt bridge/hydrogen-bond network centered on Arg44I. The side-chains of Arg44I, Glu39I and Ser41I are shown. Dashed lines represent possible hydrogen bonds or salt bridges within 3.3 Å.

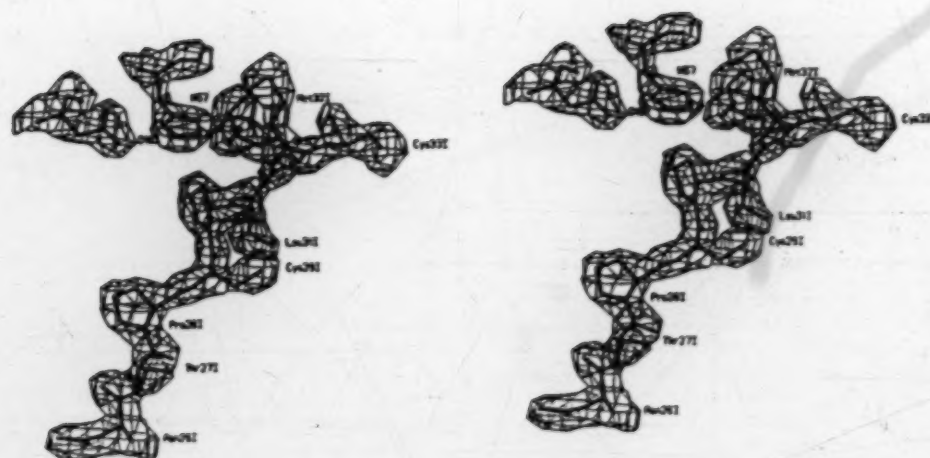


Figure 2.8 Electron density in the active site of the complex, showing the catalytic triad His57, Asp102 and Ser195 of elastase and residues Asn261 to Cys331 of the reactive-site loop of the C/E-1 inhibitor. The map was calculated with coefficients $(2m|F_o| - d|F_c|)$ and calculated phases (Read, 1986), contoured at 2.0σ level, and was displayed on a Silicon Graphics workstation using a locally modified version of the program FRODO (Jones, 1985).

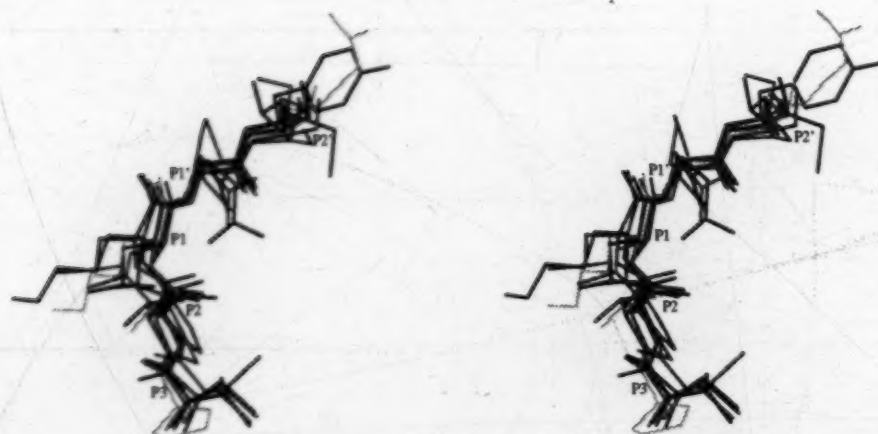


Figure 2.9 Superposition of residues P_3 to P_2' of the reactive-site loops from a selection of protein inhibitors of serine proteinases. The C/E-1 inhibitor is shown in green, bovine pancreatic trypsin inhibitor (BPTI) in yellow, ovomucoid inhibitor third domain from turkey (OMTKY3) in red, chymotrypsin inhibitor-1 from potato (PCI-1) in cyan, leech inhibitor eglin-C in purple and Bowman-Birk inhibitor from beans in pink.

Table 2.1

Comparisons of residues in C/E-1 from P3 to P2' of the reactive site loop with equivalent residues in a selection of protein inhibitors of serine proteinases.

Inhibitor	P3	P2	P1	P1'	P2'	rms difference of the 20 main-chain atoms
C/E-1 ^a	C	P	L ³¹¹	M	C	
BPTI ^b	P	C	K ¹⁵¹	A	R	0.87 Å
OMTKY3 ^c	C	T	L ¹⁸¹	E	Y	0.89 Å
PCI-1 ^d	C	P	L ³⁸¹	N	C	0.76 Å
EG-C ^e	V	T	L ⁵⁹¹	D	L	0.74 Å
B-Birk ^f	C	T	K ²⁶¹	S	M	0.73 Å

^aThe C/E-1 inhibitor from *Ascaris suum*.

^bBovine pancreatic trypsin inhibitor.

^cOvomucoid inhibitor third domain from turkey.

^dChymotrypsin inhibitor-1 from potato.

^eLeech inhibitor eglin c.

^fBowman-Birk inhibitor from Adzuki beans.

The distance from the carbonyl-carbon of Leu31I (P_1 residue) to O γ of Ser195 is 2.9Å in one C/E-1 inhibitor and 3.1Å in the other C/E-1 inhibitor in the asymmetric unit. These distances are 0.2-0.4Å longer than the unusually short (2.7 ± 0.1 Å) non-bonded contact distances observed in other serine proteinase-inhibitor complexes, including OMTKY3-SGPB (Read et al., 1983), OMTKY3-Chymotrypsin (Fujinaga et al., 1987), OMTKY3-HLE (Bode et al., 1986), PTI-BT (Marquart et al., 1983) and PCI-1-SGPB (Greenblatt et al., 1989), etc. These longer distances in the C/E-1 complexes may reflect a less-than-ideal complementarity between the reactive-site loop of the C/E-1 inhibitor and the active site of elastase or the relatively limited resolution (2.4Å) of the present determination.

Enzyme-Inhibitor Complex

A unique feature of the C/E-1 inhibitor and elastase complex is their mutual penetration. This is characterized by the penetration of the P_1 residue, Leu31I, into the S1 substrate specificity pocket, and by Arg217A (Arg217A is an insertion in the elastase structure relative to the sequence of chymotrypsin) from elastase penetrating through a pore in the C/E-1 inhibitor (Figure 8). The latter penetration has not been observed in any other protein inhibitor-proteinase complex. The pore in the C/E-1 inhibitor is about 5Å in diameter, encircled by the long loop enclosing the disulfide bridge of Cys17I to Cys29I. The side-chain of Arg217A inserts into the pore, and in so doing it makes a total of 14 interactions with residues lining the pore.



Figure 2.10 A schematic representation of the penetration of Arg217A of elastase through the pore of the C/E-1 inhibitor. The van der Waals surface of elastase is shown in white. Arg217A, whose van der Waals surface is finger-shaped, projects from the main-body of elastase. The backbone of the C/E-1 inhibitor is depicted as a red tube. The five disulfide bridges are shown in yellow. The side-chains of Leu31I and Met32I are shown in green and orange, respectively. (Diagram generated using GRASP, Nicholls et al. 1991).

Consequently the side chain atoms of Arg217A are better ordered in this complex than they are in the native elastase, as they display significantly lower B-factors in the complex. This pore is probably preformed, since the free C/E-1 inhibitor, as shown by the NMR study on the related trypsin inhibitor (ATI, Gronenborn et al., 1990), displays a very similar (albeit limited) secondary structure to the molecular complexes in this crystal structure. On the other hand, association with a cognate enzyme can induce conformational changes in the inhibitor, as in the case of chymotrypsin inhibitor-2 from barley seeds (McPhalen & James 1987). Since chymotrypsin does not have the residue corresponding to Arg217A, the structure of the complex of chymotrypsin with the C/E-1 inhibitor should help to clarify this point.

The intermolecular interactions between the C/E-1 inhibitor and elastase are similar to those observed in other complexes of protein inhibitors with their cognate proteinases. The C/E-1 inhibitor has secondary binding subsites elongated on both sides of the scissile bond. In total, 105 contacts (within 4 Å) are present between the C/E-1 inhibitor and elastase (see Table 2.2). Equal numbers of residues (20 from elastase and 20 from the C/E-1 inhibitor) are involved in these enzyme-inhibitor interactions. By comparison, only 11 residues of ovomucoid inhibitor third domain from turkey (OMTKY3) are involved in interactions in OMTKY3-*Streptomyces griseus* protease B (SGPB) complex (Read et al., 1983), and 10 residues of OMTKY3 are involved in the OMTKY3-chymotrypsin complex (Fujinaga et al., 1987). Also, 10 residues of polypeptide chymotrypsin inhibitor-1 (PCI-1)

Table 2.2

Interactions (within 4.0 Å) made between the C/E-1
inhibitor (listed vertically) and PPE.

Residues	T4	H57	R61	N148	W172	T175	G190	C191	Q192	G193	D194	S195	T213	S214	F215	V216	S217	R217A	L218	K224	Sum
P16									1												1
P15																		(1)			1
P14																		(1)			1
P13																		1			2
P7												1					1				2
P6																					14
P5																					8
P4																					6
P3																					14
P2																					5
P1																					29
P10																					6
P2																					2
P3																					1
P4																					3
P6																					1
P10																					2
P11																					2
P12																					2
P11																					3
Sum	3	3	3	3	6	9	5	2	4	9	4	1	10	2	5	6	12	4	14	2	105

The interactions listed in this table are observed in one copy of the complex. Both copies of the complex exhibit minor differences. The number of interactions may change slightly when refinement with higher resolution data is completed. The numbers in parentheses are the numbers of hydrogen bonds included in the total.

are involved in the PCI-1-SGPB complex (Greenblatt et al., 1989). Apart from those residues interacting solely with Arg217A from elastase, residues at positions P_7 , P_{18} , P_8' and P_{13}' of the C/E-1 inhibitor are also involved in intermolecular contacts in the complex. These intermolecular interactions were not observed in any of the previously determined complex structures.

Comparison of Elastase in the Free and Complexed States

The structure of elastase in the complex agrees closely with that of elastase in the native, uncomplexed state (Meyer et al., 1988). Pairwise superpositions of the native elastase structure with this complexed elastase show that 212 out of the total 240 C^α atoms fit within 1.0 Å, while 28 C^α atoms deviate greater than 1.0 Å. The rms deviation for the 212 C^α atoms is 0.56 Å and 0.55 Å, respectively for the two copies of elastase in the complex. The two copies of elastase in the asymmetric unit are more similar, having an rms deviation of only 0.35 Å for all main-chain atoms. Almost all the 28 residues exhibiting deviations greater than 1.0 Å are located on surface loops, except for two of the residues on one of the loops in the substrate binding pocket. Most of these displacements may be attributed to the difference in crystal packing environment between the native crystal and the complex crystal, and to the binding of the inhibitor to elastase. Close examination of each displacement showed that indeed the intermolecular-interaction pattern in that region is different in the complex relative to that in the native state. A

significant change takes place in the calcium-binding loop (Asn74 to Gly78). Consequently, the expected six-coordination geometry for Ca^{2+} binding is disrupted. In the complex, there is a five-coordinated water molecule occupying the expected calcium-binding site of native elastase. The corresponding loop (Ser75 to Lys79) in the chymotrypsin-OMTKY3 complex also undergoes considerable shifts in respect to the native chymotrypsin. There is a minor conformational readjustment in the polypeptide chain bordering the elastase active site (Cys191 to Gln192). Residues Cys191 to Gln192 have moved by an average of 1.2Å relative to the native elastase, presumably to allow favorable contacts with the bound C/E-1 inhibitor at Leu311.

Chemical Determination of Disulfide Bond Pairing

To learn as much as possible about the structure of *Ascaris* serine proteinase inhibitors, Bernard and Peanasky in South Dakota began a study to assign the pairing of the 10 half-cystine residues by chemical methods (Bernard and Peanasky, 1993). Native inhibitors were found to be resistant to both chemical and enzymatic fragmentation. Collaborations with laboratories in Edmonton (X-ray crystallography) and at the NIH (NMR spectroscopy) were initiated. Samples of native *Ascaris suum* serine proteinase inhibitors were made available; the same preparations were sent to both laboratories. In our laboratory we studied the C/E-1 inhibitor; the NMR group studied *Ascaris* trypsin inhibitor-1 (ATI) (Gronenborn et al., 1990).

Treatment of methionine-containing inhibitors (C/E-inhibitors) with cyanogen bromide inactivated the inhibitors, cleaved the molecules at expected peptide bonds, but did not release cystine-containing peptides. A variety of proteinases were unable to release cystine containing peptides even though they were able to hydrolyze some peptide bonds. Cleavage with cyanogen bromide was then combined with hydrolysis by glycyl endopeptidase (Gly-C) from the latex of *Carica papaya* and followed by hydrolysis by *Staphylococcal* serine proteinase (Glu-C). All three steps, employed in sequencing, were required to release cystine containing peptides. The peptides were resolved by HPLC, repurified, analyzed for amino acid composition and then sequenced using a manual double coupling method (Chang, 1981). When this study was completed, it was submitted for publication. Within weeks after the chemical study was published, the X-ray laboratory solved the C/E-1 structure and made an assignment of the disulfide bonds. Regrettably, the published disulfide bond assignment made using chemical methods supports neither the structure determined by X-ray nor disulfide bond assignments made by calculations from the NMR-spectra of the *Ascaris* trypsin inhibitor [M. Clore, personal communication]. The disulfide bond assignments by X-ray analysis and by calculations from the NMR spectra agree with each other.

The experiments and data on which the chemically determined disulfide bond assignments were based were reviewed. It appears that the fragmentation data obtained permit the assignments reported by Bernard and Peanasky (1993). There is no clear explanation for the discrepancy. It is possible that the inactivation with cyanogen

bromide and the unique arrangement of the five disulfide bonds in native inhibitors may have permitted the disulfides to rearrange either to some intermediate folding form or a dead-end product which was unwittingly determined.

Biological Implications

It has been proposed (Laskowski & Kato, 1980) that most protein inhibitors of serine proteinases act by a common mechanism; they bind very tightly to their cognate enzymes, but are hydrolyzed very slowly. This tight binding has been explained by crystallographic studies, which have revealed extensive surface complementarity between the reactive-site loops of the inhibitors and the active-site clefts of the enzymes. Thus, the protein inhibitors recognize and bind to their cognate enzymes in the manner of a good substrate. Presumably, the excellent fit throughout the enzyme-inhibitor interface (marked by a large area of contact of 1200-1800 Å², as calculated with the program MS of Connolly, 1983) results in a remarkably deep minimum in the free-energy profile, thereby presenting a very high activation energy barrier to hydrolysis and to dissociation. In the complex of elastase with the C/E-1 inhibitor, the complementarity is extremely good throughout the reactive-site loop as well as for more peripheral regions (the area of contact between the enzyme and the inhibitor is ~2100 Å²). For example, the side chain of Arg217A of elastase fits neatly into a preformed pore in the inhibitor, without any collision. The reactive-site loop of the C/E-1 inhibitor is similar in conformation to those of most typical proteinase inhibitors,

indicating that the C/E-1 inhibitor probably conforms to the standard inhibitory mechanism.

The structure of the C/E-1 inhibitor presented here confirms that this inhibitor indeed differs from the other known inhibitor families in tertiary structure and topography of disulfide bridges, and thus represents a novel family of inhibitors. On the basis of the high sequence identity and the conserved positions of cysteine residues, it is likely that the other seven known serine proteinase inhibitors of the *Ascaris* family will also adopt a structure similar to that of the C/E-1 inhibitor.

Hydrophobic interactions are believed to be the main driving forces in the protein folding process. The absence of a hydrophobic core in the C/E-1 inhibitor is an exception to the general rule, thereby challenging the current folding algorithms. The five disulfide bridges, as well as the elaborate network of electrostatic and hydrogen-bonding interactions, appear to be the major stabilizing forces in the C/E-1 inhibitor. Mutagenesis of each disulfide pair would readily reveal its significance in the folding process, and in maintaining the structure and function of the C/E-1 inhibitor. An analogous observation was made with an acid-stable human mucous proteinase inhibitor (SLPI) (Grutter et al., 1988), which is also rich in disulfides and lacks extensive secondary structure. Buried charged residues, other than hydrophobic residues, comprise the molecular core of SLPI. These small proteins may provide valuable model systems with

which to study how the primary sequence of a protein dictates its tertiary structure.

MATERIALS and METHODS

Crystallization of the Complex

Porcine pancreatic elastase (Lot 31013C; Serva Feinbiochemica GmbH and Co.) was used without further purification (Lot. 31013C). The C/E-1 inhibitor was purified as described previously (Peanasky et al., 1984). The complex was prepared by mixing the C/E-1 inhibitor with elastase in a 1 to 1.1 molar ratio of enzyme to inhibitor. Crystals of the complex were grown from solutions of 10 mg/ml protein and 12% PEG 6000, buffered with 50 mM sodium citrate at pH 6.5. Two crystal forms were observed from growth at different temperatures. Extremely thin plate-shaped crystals (dimensions $0.6 \times 0.2 \times 0.001 \text{ mm}^3$) were obtained at room temperature; they belong to space group $P2_12_12$, with unit cell parameters $a = 70.45\text{\AA}$, $b = 114.33\text{\AA}$, $c = 74.55\text{\AA}$. Growth at 4°C yielded large chunky crystals (dimensions $0.4 \times 0.3 \times 0.2 \text{ mm}^3$) which belong to space group $P3_221$, with $a = b = 84.02\text{\AA}$, $c = 190.93\text{\AA}$. Both crystal forms have values of V_m (Matthews, 1968) that suggest two sets of molecular complexes per asymmetric unit.

Data collection

The trigonal crystal form was chosen for the present structure determination because of its better diffracting ability. Crystals

were mounted into glass capillaries at 4°C. Diffraction data were recorded at room temperature, with a Siemens area detector, X-1000, mounted on an 18 kW Siemens rotating anode X-ray generator. Graphite monochromatized Cu K α radiation was used for data collection. Three data sets were collected from three crystals; each is nearly complete to 2.4Å. All three data sets were merged and used for the structure solution and refinement. The data processing and merging were performed with the Xengen package (Howard et al., 1987). Data collection statistics are given in Table 2.3.

Structure determination

The initial phases were determined by the molecular replacement method, using the native porcine pancreatic elastase (Meyer et al., 1988) as the search model. Reflections between 20-3.5Å were used for the rotation function search that was conducted with the program ROTING (Fujinaga & Read, 1987). The highest peak given by the Navaza rotation function search was 8.5 σ above the mean; the second highest peak was significantly lower (4.2 σ). This suggested that the two sets of complex molecules were in nearly the same orientation in the asymmetric unit. The translation search was done using BRUTE (Fujinaga & Read, 1987), with the two elastase molecules in an identical orientation in the unit cell. It gave two peaks, one for each elastase molecule. Rigid-body refinement in X-PLOR (Brunger et al., 1987) was used to optimize the molecular replacement solution. The R-factor that resulted was 41.8%, the correlation coefficient was 54.3%, with two elastase molecules in one asymmetric unit.

Table 2.3
Data collection and refinement

Space group	P3 ₂ 21
Cell dimensions	a=b=84.02Å, c=190.93Å
Data collection	
Maximum resolution	2.4Å
No. of crystals used	3
Total observations	192,689
Unique reflections	29,261
Average redundancy	6.58
Completeness of data	20Å-2.5Å: 99% 2.5Å-2.4Å: 71%
R _{merge} ^a	7.00%
Refinement statistics	
No. of protein atoms	4538
No. of solvent atoms	100
Missing residues	I62, I63
Reflections used	29,261
Resolution range	20Å-2.4Å
R-factor	19.1%
rms deviation from ideal	
bond distance	0.021Å
bond angle	3.8°
planar groups	0.026Å

$$^a R_{\text{merge}} = \sum_{hkl} \left(\frac{\sum_i |I_i - \langle I \rangle|}{\sum_i I_i} \right)$$

An initial 2.4 Å electron density map computed after the rigid-body refinement showed very good electron density for the two elastase molecules. So the remaining task was to build the C/E-1 inhibitor into the complex. The presence of two sets of molecular complexes per asymmetric unit made electron density averaging feasible. The DEMON package was used for this purpose (Vellieux et al., 1993). A local correlation map was calculated to generate the initial molecular mask embracing a single copy of the complex. This defined the region where averaging should take place. After several cycles of two-fold averaging and solvent flattening, 85% of the residues of the C/E-1 inhibitor could be built into the resulting electron density map. The molecular envelope was then adjusted accordingly for further averaging cycles. Eventually 61 residues of the C/E-1 inhibitor could be built into the density unambiguously. The last two residues at the carboxyl terminus, Glu62I and His63I, were not well defined in this or subsequent electron density maps.

Refinement

The complex model was subjected to energy minimization and simulated annealing molecular dynamics refinement using the X-PLOR package. Reflections between 10-2.4 Å with intensities greater than $3\sigma(I)$ were used in the X-PLOR refinement. The model was then refined with the program TNT (Tronrud, 1992), using all reflections between 20-2.4 Å with no α cut-off. The initial B-factors of the elastase residues in the complex were taken from the native elastase structure (Meyer et al., 1988); the initial B-factors of the C/E-1 inhibitor

residues were assigned as $15.0\text{\AA}^2$. The positional parameters and the B-factors were refined at the same time while the occupancy factors were kept at 1.0 for all atoms. About 50 ordered solvent molecules have been added to each complex. The current model contains two copies of elastase, two copies of the C/E-1 inhibitor (from which the last two residues are missing), and 100 solvent molecules. The present quality of the complex structure is summarized in Table 2.3.

REFERENCES

- Abuereish GM & Peanasky RJ. 1974. Pepsin inhibitors from *Ascaris lumbricoides*: Isolation, purification and some properties. *J Biol Chem* 249:1558-1565.
- Babin DR, Peanasky RJ & Goss SM. 1984. The isoinhibitors of chymotrypsin/elastase from *Ascaris lumbricoides*: the primary structure. *Arch Biochem Biophys* 232:143-161.
- Baumann U, Huber R, Bode W, Grosse D, Lesjak M & Laurell CB. 1991. Crystal structure of cleaved human α -1-antichymotrypsin at 2.7Å resolution and its comparison with other serpins. *J Mol Biol* 218:595-606.
- Bernard VD & Peanasky RJ. 1993. The serine protease inhibitor family from *Ascaris suum*: Chemical determination of the five disulfide bridges. *Arch Biochem Biophys* 303:367-376.
- Bigler TL, Lu W, Park SJ, Tashiro M, Wieczorek M, Wynn R & Laskowski M. 1993. Binding of amino acid side chains to preformed cavities: Interaction of serine proteinases with turkey ovomucoid third domains with coded and noncoded P1 residues. *Protein Science* 2:786-799.
- Bode W, Epp O, Huber R, Laskowski M Jr & Ardelt W. 1985. The crystal and molecular structure of the 3rd domain of silver pheasant ovomucoid (OMSVP3). *Eur J Biochem* 147:387-395.
- Bode W, Wei AZ, Huber R, Meyer E, Travis J & Neumann S. 1986. X-ray crystallographic structure of the complex of human leukocyte elastase (PMN elastase) and the 3rd domain of the turkey ovomucoid inhibitor. *EMBO J*. 5:2453-2458.

- Bode W, Greyling HJ, Huber R, Otlewski J & Wilusz T. 1989a. The refined 2.0Å X-ray crystal structure of the complex formed between bovine β -trypsin and CMTI-I, a trypsin inhibitor from squash seeds (*Curcubita maxima*): Topological similarity of the squash seed inhibitors with the carboxy peptidase A inhibitor from potatoes. *FEBS Lett* 242:285-292.
- Bode W, Meyer E & Powers J. 1989b. Human leukocyte and porcine pancreatic elastase: X-ray crystal structures, mechanism, substrate specificity and mechanism-based inhibitors. *Biochem* 28:1951-1963.
- Bode W & Huber R. 1992. Natural protein proteinase inhibitors and their interaction with proteinases. *Eur J Biochem* 204:433-451.
- Bolognesi M, Pugliese L, Gatti G, Frigerio F, Coda A, Antolini L, Schnebli HP, Menegatti E, Amiconi G & Ascenzi P. 1990. X-ray crystal structure of the bovine α -chymotrypsin/eglin c complex at 2.6 Å resolution. *J Mol Recog* 3:163-168.
- Brünger AT, Kuriyan J & Karplus M. 1987. Crystallographic R factor refinement by molecular dynamics. *Science* 235:458-460.
- Carrell RW, Stein PE, Fermi G & Wardell M. 1994. Biological implications of a 3Å structure of dimeric antithrombin. *Structure* 2:257-270.
- Chang JY. 1981. N-terminal sequence analysis of polypeptides at the picomole level. *Biochem J* 199:557-564.
- Chen, P., Rose, J., Love, R., Wei, C. H. & Wang, B. C. 1992. Reactive sites of an anticarcinogenic Bowman-Birk proteinase inhibitor are similar to other trypsin inhibitors. *J Biol Chem* 267:1990-1994.
- Connolly ML. 1983, Analytical molecular surface calculation. *J Appl Cryst* 16:548-558.

- Dauter Z & Betzel C 1991. Complex between the subtilisin from a mesophilic bacterium and the leech inhibitor eglin-C. *Acta Cryst B* 47:707-730.
- Delarue M, Samama JP. & Moras D. 1990. Crystal structure of bovine antithrombin III. *Acta Cryst B* 46:550-556.
- Fujinaga M, Sielecki AR, Read RJ., Ardelt W, Laskowski M Jr, & James MNG. 1987. The crystal and molecular structures of the complex of turkey ovomucoid inhibitor third domain with α -chymotrypsin at 1.8 Å resolution. *J Mol Biol* 195:397-418.
- Fujinaga M & Read RJ. 1987. Experiences with a new translation-function program. *J Appl Cryst* 20:517-521.
- Goodman RB & Peanasky RJ. 1982. Isolation of the trypsin inhibitors in *Ascaris Lumbricoides* var suum using affinity chromatography. *Anal Biochem* 120:387-393.
- Greenblatt HM, Ryan CA & James MNG. 1989. Structure of the complex of *Streptomyces griseus* proteinase B and polypeptide chymotrypsin inhibitor-1 from Russet Burbank potato tubers at 2.1Å. *J Mol Biol* 205:201-228.
- Gronenborn AM, Nilges M, Peanasky RJ & Clore GM. 1990. Sequential resonance assignment and secondary structure determination of the *Ascaris* trypsin inhibitor, a member of a novel class of proteinase inhibitors. *Biochemistry* 29:183-189.
- Grutter MG, Fendrich G, Huber R & Bode W. 1988. The 2.5 Å X-ray crystal structure of the acid-stable proteinase inhibitor from human mucous secretions analyzed in complex with bovine α -chymotrypsin. *EMBO J* 7:345-351.
- Grutter MG, Priestle JP, Rahuel J, Grossenbacher H, Bode W, Hofsteenge J & Stone SR. 1990. Crystal structure of the thrombin-hirudin

complex: a novel mode of serine proteinase inhibition. *EMBO J* 9:2361-2365.

Hirono S, Akagawa H, Mitsui Y & Iitaka Y. 1984. Crystal structure at 2.6Å resolution of the complex of subtilisin BPN' with *Streptomyces* subtilisin inhibitor. *J Mol Biol* 178:389-413.

Homandberg GA & Peanasky RJ. 1976. Characterization of proteins from *Ascaris Lumbricoides* which bind specifically to carboxy peptidase. *J Biol Chem* 251:2226-2233.

Howard AJ, Gilliland GL, Finzel BC. & Poulos TL. 1987. The use of an imaging proportional counter in macromolecular crystallography. *J Appl Crystallogr* 20:383-387.

Jones TA. 1985. Interactive, computer graphics: FRODO. With modifications for TOM UQV/CIT v.2.8.0 by C. Cambillau, M. Israel, G. Griffin and A. Chirino. *Meth Enzym* 115:157-171.

Kabsch W & Sander C. 1983. Dictionary of protein secondary structure: pattern recognition of hydrogen bonded and geometrical features. *Biopolymers* 22:2577-2637.

Kraulis PJ. 1991. MOLSCRIPT: a program to produce both detailed and schematic plots of protein structures. *J Appl Cryst* 24:946-950.

Laskowski M Jr & Kato I. 1980. Protein inhibitors of proteinases. *Annu Rev Biochem* 49:593-626.

Lin G, Bode W, Huber R, Chen C. & Engh RA. 1993. The 0.25-nm X-ray structure of the Bowman-Birk type inhibitor from Mung bean in ternary complex with porcine trypsin. *Eur J Biochem* 212:549-555.

Lobermann H, Tokuoka R, Deisenhofer J & Huber R. 1984. Human α 1-proteinase inhibitor: crystal structure analysis of two crystal modifications, molecular model and preliminary analysis of the implications for function. *J Mol Biol* 177:531-556.

- Marquart M, Walter J, Deisenhofer J, Bode W & Huber R. 1983. The geometry of the reactive site and of the peptide groups in trypsin, trypsinogen and its complex with inhibitors. *Acta Cryst B39*:480-490.
- Martzen MR, Geise GL, Hogan BJ. & Peanasky RJ. 1985. *Ascaris suum*: Localization by immunochemical and fluorescent probes of host proteases and parasite proteinase inhibitors in cross-sections. *Expl Parasitol* 60:139-149.
- Matthews BW. 1968. Solvent content of protein crystals. *J Mol Biol* 33:491-497.
- McPhalen CA, Svendsen I, Jonassen I & James MNG. 1985a. Crystal and molecular structure of the inhibitor CI-2 from barley seeds in complex with subtilisin Novo. *Proc Natl Acad Sci USA* 82:7242-7246.
- McPhalen CA, Schnebli HP & James MNG. 1985b. Crystal and molecular structure of the inhibitor eglin from leeches in complex with subtilisin Carlsberg. *FEBS Lett* 188:55-58.
- McPhalen CA & James MNG. 1987. Crystal and molecular structure of the serine proteinase inhibitor CI-2 from barley seeds. *Biochemistry* 26:261-269.
- McPhalen CA & James MNG. 1988. Structural comparison of two serine proteinase-protein inhibitor complexes: Eglin-C-subtilisin Carlsberg and CI-2-subtilisin Novo. *Biochemistry* 27:6582-6598.
- Meyer E, Cole G & Radharkrishnan R. 1988. Structure of native porcine pancreatic elastase at 1.65 Å resolution. *Acta Cryst B44*:26-38.
- Navaza J. 1990. Accurate computation of the rotation matrices. *Acta Cryst A46*:619-620.
- Nicholls A Sharp KA & Honig B. 1991. Protein folding and association: Insights from the interfacial and thermodynamic properties of hydrocarbons. *Proteins: Struct Funct and Genet* 11:281-296.

- Onesti S, Brick P & Blow DM. 1991. Crystal structure of a Kunitz-type trypsin inhibitor from *Erythina caffra* seeds. *J Mol Biol* 217:153-176.
- Papamokos E, Weber E, Bode W, Huber R, Empie MW, Kato I & Laskowski M Jr. 1982. Crystallographic refinement of Japanese quail ovomucoid, a Kazal-type inhibitor, and model building studies of complexes with serine proteases. *J Mol Biol* 158:515-537.
- Peanasky RJ, Bentz Y, Paulson B, Graham DL & Babin DR. 1984a. The isoinhibitors of chymotrypsin/elastase from *Ascaris Lumbricoides*: Isolation by affinity chromatography and association with the enzymes. *Arch Biochem Biophys* 232:127-134.
- Peanasky RJ, Bentz Y, Homandberg GA, Minor ST & Babin DR. 1984b. The inhibitors of chymotrypsin/elastase from *Ascaris Lumbricoides*: The reactive site. *Arch Biochem Biophys* 232:135-142.
- Peanasky RJ, Martzen MR, Homandberg GA, Cash JM, Babin DR & Litweiler B. 1987. Proteinase inhibitors from intestinal parasitic helminths: Structure and indications of some possible functions. In *Molecular paradigms for eradicating helminthic parasites*. MacInnis AJ, eds. New York:DPL. pp. 349-366.
- Read RJ, Fujinaga M, Sielecki AR & James MNG. 1983. Structure of the complex of *Streptomyces griseus* protease B and the third domain of the turkey ovomucoid inhibitor at 1.8 Å resolution. *Biochemistry* 22:4420-4433.
- Read RJ & James MNG. 1986. Introduction to the protein inhibitors: X-ray crystallography. In *Proteinase Inhibitors* Barrett AJ & Salvesen G, eds, Amsterdam:Elsevier. pp. 301-336,
- Read RJ. 1986. Improved Fourier coefficients for maps using phases from partial structures with errors. *Acta Cryst* A42:140-149.

- Rydel TJ, Ravichandran KG, Tulinsky A, Bode W, Huber R, Roitsch C & Fenton JW. 1990. The structure of a complex of recombinant hirudin and human α -thrombin. *Science* 249:277-280.
- Schechter I & Berger A. 1967. On the size of the active site in proteases. I. Papain. *Biochem Biophys Res Commun* 27:157-162.
- Schreuder HA, De Boer B, Dijkema R, Mulders J, Theunissen HJM, Grootenhuis PDJ & Hol WG. 1994. The intact and cleaved human antithrombin III complex as a model for serpin-proteinase interactions. *Nature Struct Biol* 1:48-54.
- Shotton DM & Watson HC. 1970. Three-dimensional structure of tosyl-elastase. *Nature* 225:811-816.
- Stein PE, Leslie AGW, Finch JT, Turnell WG, McLaughlin PJ & Carrell RW. 1990. Crystal structure of ovalbumin as a model for the reactive centre of serpins. *Nature* 347:99-102.
- Sweet RM, Wright HT, Janin S, Chothia CH & Blow DM. 1974. Crystal structure of the complex of porcine trypsin with soybean trypsin inhibitor (Kunitz) at 2.6Å resolution. *Biochemistry* 13:4212-4228.
- Tronrud DE. 1992. Conjugate-direction minimization. An improved method for the refinement of macromolecules. *Acta Cryst A* 48:912-916.
- Tshuogae Y, Tanaka I, Yamane T, Kikkawa J, Achida JT, Ishikawa C, Watanabe K, Nakamura S & Takahashi K. 1986. Structure of the trypsin binding domain of Bowman-Birk type proteinase inhibitor and its interaction with trypsin. *J Biochem* 100:1637-1646.
- Vellieux FMD, Hajdu J, Verlinde LMJ, Groendijk H, Read RJ, Greenhough TJ, Campbell JW, Kalk KH, Littlechild JA, Watson HC & Hol WGJ. 1993. Structure of glycosomal glyceraldehyde-3-phosphate dehydrogenase from *Trypanosoma brucei* determined from Laue data. *Proc Natl Acad Sci USA* 90:2355-2359.

Wlodawer A, Deisenhofer J & Huber R. 1987. Comparison of two highly refined structures of bovine pancreatic trypsin inhibitor. *J Mol Biol* 193:145-156.

Wright HT, Qian HX & Huber R. 1990. Crystal structure of Plakalbumin, a proteolytically nicked form of ovalbumin. Its relationship to the structure of cleaved α 1-proteinase inhibitor. *J Mol Biol* 213:513-528.

CHAPTER 3

Water Molecules Participate in Proteinase-
Inhibitor Interactions: Crystal Structures of the
Leu¹⁸, Ala¹⁸ and Gly¹⁸ Variants of Turkey Ovomucoid
Inhibitor Third Domain Complexed With *Streptomyces*
Griseus Proteinase B*

INTRODUCTION

In recent years many proteins from diverse sources have been shown to be inhibitors of a variety of proteinases. Thus far, twenty or so families of protein-proteinase inhibitors have been identified (reviewed by Bode & Huber, 1992). Proteinase inhibitors are classified into different families on the basis of their target enzymes, their inactivation mechanisms, their primary sequences and their tertiary structures. Protein inhibitors of the serine proteinases exhibit a standard mechanism of inhibition (with the serpin and hirudin families as the only known exceptions), despite their unrelated sequences and three-dimensional structures (Laskowski & Kato, 1980). These so-called canonical inhibitors bind to their cognate enzymes in the same manner as a good substrate but are cleaved extremely slowly. The inhibitors resemble good substrates in two aspects: they display extremely tight binding (with K_d up to 10^{12} M^{-1}) and have very strict specificity (as

*A version of this chapter has been accepted for publication. Huang et al., 1995. *Protein Science*.

measured by k_{cat}/K_m) for their cognate enzymes. X-ray structures of these substrate-like inhibitors have revealed that the reactive-site loops of the inhibitors are chemically and geometrically complementary to the active sites of their cognate enzymes and that the complexes are non-covalent (reviewed by Read & James, 1986). These intrinsic structural features of protein inhibitors confer them with exceptionally high binding affinities for their cognate enzymes.

The primary specificity determinant of a given inhibitor is the residue at the P_1 position, the residue contributing the carbonyl portion to the reactive-site peptide bond (nomenclature of Schechter & Berger, 1967). Substitution of the P_1 residue results in a dramatic change in K_i , by six orders of magnitude in some cases, corresponding to free energies of binding of up to 8 Kcal/mol, as displayed by the variants of the ovomucoid inhibitor third domain from turkey (Bigler et al., 1993; Lu, 1994). OMTKY3 is one of the best-studied protein inhibitors. Native OMTKY3 has 56 amino acids; the peptide bond between Leu¹⁸ and Glu¹⁹ is the scissile bond and thus is termed the reactive-site bond. OMTKY3 is a potent inhibitor of chymotrypsin ($K_i=1.9 \times 10^{11} \text{ M}^{-1}$). When the P_1 residue, Leu¹⁸, of OMTKY3 is replaced by lysine, the resulting variant becomes a strong inhibitor of trypsin. Similarly, the Glu¹⁸ variant is a strong inhibitor of the Glu-specific SGP (*Streptomyces griseus* proteinase) (Komiya et al, 1991). The Arg¹⁵-Lys¹⁷-Arg¹⁸ variant of OMTKY3 is a good inhibitor of human furin (Lu et al, 1993).

A wealth of kinetic and thermodynamic data makes OMTKY3 an attractive model to study the inhibitory mechanism used by protein

inhibitors. An overexpression system for OMTKY3 has been established (Lu, et al, 1993). This facilitates the generation and purification of various mutants of interest. So far the P_1 residue, which is Leu18 in the naturally occurring OMTKY3, has been substituted by the other 19 coded amino acids and 5 non-coded amino acids (Hse, Abu, Ape, Ahx and Ahp). The K_a (equilibrium constant for association) of all these recombinant P_1 variants were measured for six proteinases: SGPA (*Streptomyces griseus* proteinase A), SGPB (*Streptomyces griseus* proteinase B), chymotrypsin, porcine pancreatic elastase, human leukocyte elastase, and subtilisin (Bigler et al., 1993; Lu, 1994). These data have established an extensive characterization of the P_1 preference of the six well-studied enzymes. Thorough analysis and correlation of these valuable data to detailed X-ray structures will greatly extend our knowledge about proteinase-inhibitor interactions and subsequently will benefit inhibitor design.

SGPB is a chymotrypsin-like proteinase with respect to its substrate specificity. Measurements of the binding equilibrium constants have shown that leucine is most-favored at the P_1 position of OMTKY3, with a K_a of $5.6 \times 10^{10} \text{ M}^{-1}$ (Bigler, 1993; Lu, 1994). Proline is least-favored with a K_a of $3.6 \times 10^4 \text{ M}^{-1}$. In addition, the P_1 position prefers amino acids with straight and γ -branched side-chains over those of β -branched of an equal number of atoms. The S_1 binding pocket (the preformed cavity in the enzyme into which the P_1 residue inserts) of SGPB is charge-phobic, as the introduction of positively- or negatively-charged residues at the P_1 position shows deleterious effects on K_a . Interestingly, valine and lysine have completely different polarities and volumes, but display comparable K_a values and are essentially

"isofunctional" as the P_1 residues of OMTKY3. While modeling may provide insights for some of the results, most interpretations are far from quantitative and convincing.

Crystal structures of OMTKY3 have been determined in complexes with SGPB (Read et al., 1983), α -chymotrypsin (Fujinaga et al., 1987) and human leukocyte elastase (Bode et al., 1986). The crystal structures of two closely related inhibitors, ovomucoid inhibitor third domains from Japanese quail (OMJPD3) and from silver pheasant (OMSPV3), have been solved both in free, intact form (Weber et al., 1981; Papamokos et al., 1982; Bode et al., 1985) and in the scissile bond cleaved form (Musil et al., 1991). We have undertaken crystal structure determinations of all the recombinant P_1 variants of OMTKY3 (excluding non-coded amino-acid variants) in complex with SGPB. The scope of our project is to provide detailed structural data to assist in the explanation of the kinetic and thermodynamic data. One of our ultimate objectives is to be able to predict the free energy of binding for a given inhibitor and proteinase by computational simulation, and thereafter to evaluate the contribution of each individual residue involved in the complex formation. Accurate prediction of binding constants demands correct modeling of the conformational change induced by the protein-ligand association and realistic estimation of the interaction energy in that system (reviewed by Kollman, 1994). Therefore precise information concerning possible conformational changes and the solvation state of interacting residues is required and thus far can only be obtained reliably from detailed X-ray crystallographic studies.

In this study, we report the crystal structures of three OMTKY3 variants, Leu¹⁸, Ala¹⁸ and Gly¹⁸, in complex with SGPB. These variants span a large range of K_d values (4.7×10^3 fold change) in their association with SGPB ($5.6 \times 10^{10} \text{ M}^{-1}$, $3.6 \times 10^8 \text{ M}^{-1}$ & $1.2 \times 10^7 \text{ M}^{-1}$, respectively). In addition to the substitutions at the P₁ position, these recombinant variants differ from native OMTKY3 in that the first five residues of the N-terminus have been removed. Kinetic measurements have shown that the deletion of the first five residues has no effect on K_d (Wieczorek et al., 1987). The previously reported structure of native OMTKY3 in complex with SGPB also showed that the first six residues could not be located reliably in the electron density map, presumably because they were disordered or highly mobile (Read et al., 1983). We have also redetermined the structure of the recombinant Leu¹⁸ variant (residues 6 to 56) to ensure that small structural differences in the active site regions displayed by the different variants were not due to the truncation of the N-terminus. We show that there are small but important conformational changes in the active site of SGPB upon the association with the OMTKY3 variants. In adjusting to the size of the P₁ residue, water molecules bound in the S₁ pocket are redistributed, presumably to optimize the enzyme-inhibitor interactions. Water-mediated hydrogen-bonding networks, as observed in the complexes of DNA with DNA-binding proteins (Shakked, et al, 1994; Kim & Burley, 1994), peptides with MHC class I molecules (Madden, et al., 1992), and peptides with oligopeptide binding protein OppA (Tame et al., 1994), are also present in our enzyme-inhibitor complexes.

RESULTS

Quality of the complex structures

The crystallographic R-factors and the final stereochemical parameters of the current models are listed in Table 3.1. The final R-factors for the complexes of SGPB with the Leu¹⁸, Ala¹⁸, and Gly¹⁸ variants are 13.9% (1.8 Å), 17.1% (1.4 Å) and 13.2% (1.9 Å) respectively. Figure 3.1 is a representative electron density map at 1.4 Å resolution, showing the reactive-site loop and water molecules bound in the S₁ binding pocket in the complex of SGPB and the Ala¹⁸ OMTKY3 variant.

N-terminal truncation of OMTKY3 has no effect on the reactive-site conformation

All the 51 residues of the OMTKY3 variants can be located in the electron density map. Figure 3.2 shows the superposition of the main-chain atoms of native OMTKY3 (residues 1 to 56, Read et al., 1983) with the Leu¹⁸ variant of OMTKY3 (residues 6 to 56). The root-mean-square (rms) deviation for the 200 common main-chain atoms (residues 7 to 56) is 0.18 Å. There are significant shifts (greater than 2 times the rms difference) in the region of the N-terminus (Asp7 to Pro12) and the neighboring residues of the α -helix (Cys38 to Glu43) disulfide-bridged to the N-terminus. The largest shift is 0.8 Å at the N atom of residue Asp7. Elsewhere there are virtually no noticeable changes in other

Table 3.1
Statistics of structural refinement

	Leu ¹⁸ complex	Ala ¹⁸ complex	Gly ¹⁸ complex
No. of Reflections used	17,165	30,412	13,120
Resolution range (Å)	20.0 - 1.8	20.0 - 1.4	20.0 - 1.9
R-factor * (%)	13.9	17.1	13.2
No. of protein atoms	1697	1694	1693
No. of water atoms	169	182	168
No. of phosphate anions	1	1	1
rms deviation from ideal stereochemistry			
bond distance (Å)	0.022	0.023	0.023
bond angle (°)	2.4	2.4	2.5
planar groups (Å)	0.022	0.022	0.021

$$^*R\text{-factor} = \frac{\sum_{hkl} ||F_o| - |F_c||}{\sum_{hkl} |F_o|}$$

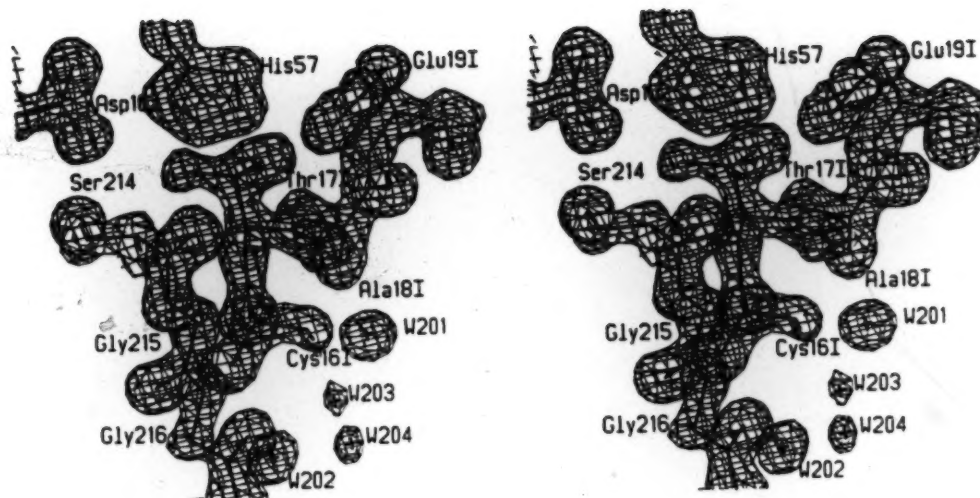


Figure 3.1. Electron density map showing part of the active site of the complex of SGPB with the Ala¹⁸ variant of OMTKY3. The map shows the catalytic His57 of SGPB, the residues flanking the scissile bond of OMTKY3 and four water molecules bound in the S₁ binding pocket. The map is computed with coefficients $(2m|F_o| - d|F_c|)\exp(i\alpha_{cal})$ (Read, 1986) and calculated phases and is contoured at the 1.5 σ level.

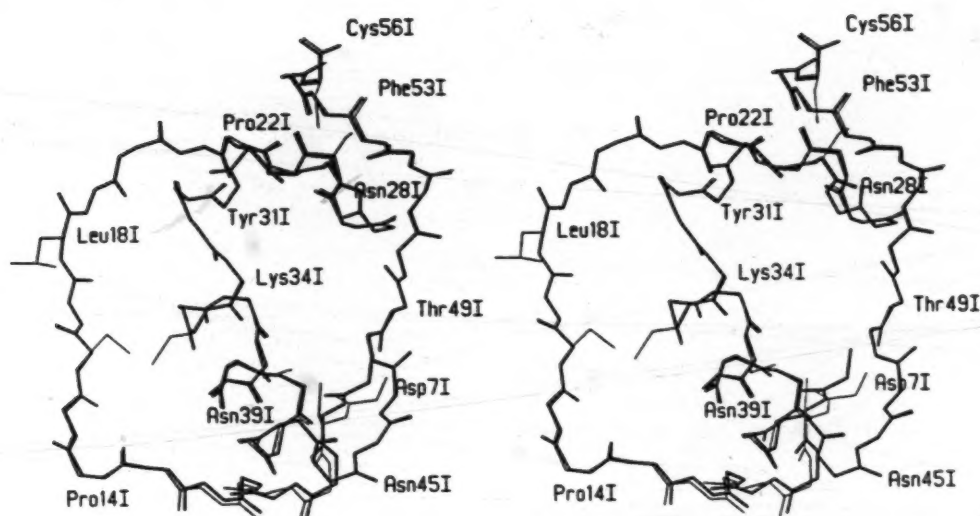


Figure 3.2. Superposition of the Leu¹⁸ variant of OMTKY3 (thick lines) with native OMTKY3 (thin lines) (Read et al., 1983). Only main-chain atoms and disulfide bridges are shown. The rms deviation for the 200 common main-chain atoms is 0.18 Å. There is no noticeable difference in the reactive-site loop.

regions including the reactive-site loop. The three variants used in this study show better agreement with one another. The rms difference of the 200 backbone atoms is 0.15 Å for Leu¹⁸ versus Ala¹⁸ variants (the largest difference is 0.52 Å at Pro14 O), 0.14 Å for Leu¹⁸ versus Gly¹⁸ variants (the largest difference is 0.50 Å at Pro14 O), and 0.12 Å for Ala¹⁸ versus Gly¹⁸ variants (the largest difference is 0.27 Å at Ser9 O). These rms differences are close to the experimental positional errors of crystallographically-determined structures at the corresponding resolutions. It is interesting to note that although the first six residues of the N-terminus are disordered in the previously reported structure, their presence, however, has an impact on the conformation of the inhibitor. Nevertheless, the conformations of the reactive-site loops are not affected by the N-terminal truncations, which is consistent with the fact that the K_i 's for native OMTKY3 and the truncated Leu¹⁸ variant of OMTKY3 are virtually identical (Wieczorek et al., 1987).

Conformational changes of one substrate binding loop of SGPB

Figure 3.3 is a main-chain atom superposition of the uncomplexed SGPB (crystallized at its enzymatically optimal pH, pH7.4, in a cubic crystal form; refined at 1.8 Å resolution) (James, et al., 1973; Blanchard & James, 1995), with SGPB in the complexed state with the Leu¹⁸ variant of OMTKY3. The rms difference for the 740 main-chain atoms of SGPB is 0.27 Å. The most significant difference between the two structures is the shift of the residues forming one side of the S₁ specificity pocket (Gly215 to Ser217). This segment of polypeptide

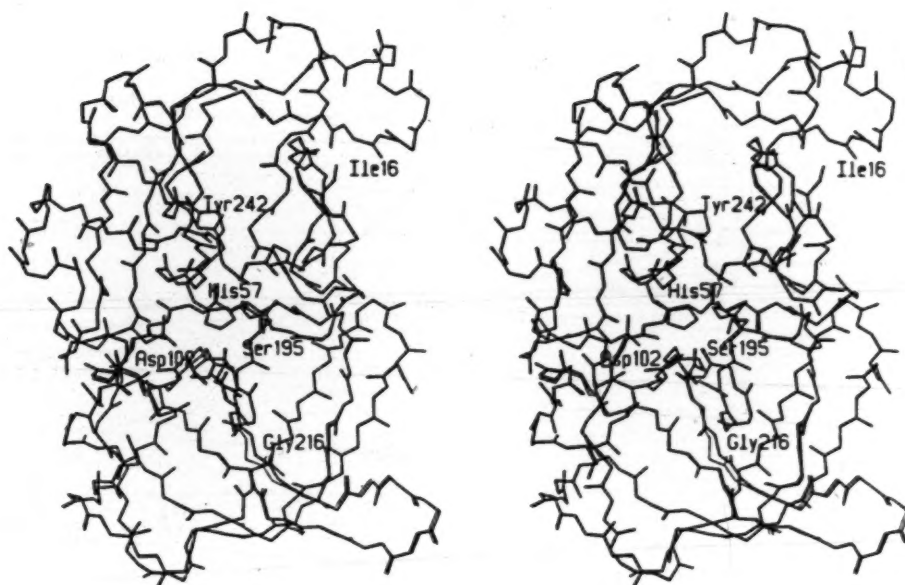


Figure 3.3 Superposition of the uncomplexed SGPB (thin lines) (Blanchard & James, 1995) with SGPB in complex with the Leu¹⁸ variant (thick lines). Residue Gly216 is labeled to highlight the conformational change of the segment Gly215 to Ser217. Gly216 C^α of SGPB moves 0.7 Å towards the side-chain of Leu18 of OMTKY3.

makes only a few direct contacts with the P_1 residue (the closest contact is 3.8 Å between Gly215 C $^{\alpha}$ of SGPB and Leu18 C $^{\delta 1}$ of OMTKY3). In response to the binding of the inhibitor, this loop moves from the native position in a concerted manner closer to the side-chain of Leu18 of OMTKY3 as if drawn by it. The C atom of Gly216 of SGPB makes the largest shifts, 0.7 Å toward the side-chain of Leu18 of OMTKY3.

It should be noted that this conformational change in SGPB takes place only in the Leu¹⁸ variant complex. There are two SGPB molecules per asymmetric unit in the uncomplexed SGPB, in them the conformations of the S_1 binding loop (Gly215 to Ser217) are virtually identical. In the complexes of SGPB with the Ala¹⁸ and Gly¹⁸ variants, loop Gly215 to Ser217 remains in the same position as in the uncomplexed SGPB.

A phosphate binding site in the interaction region

A difference map (coefficients $(|F_o| - |F_c|)\exp(i\alpha_{cal})$, where $|F_o|$ and $|F_c|$ are respectively the observed and calculated structure factor amplitudes, and α_{cal} the calculated phases) unambiguously indicates that a phosphate binding site is present (Fig. 3.4). It is identified in all the three complexes, in the peripheral region of the complex interface, at a position equivalent to W32 in the previously determined structure (Read et al, 1983). The phosphate is refined with a half occupancy and B factors around 40 Å². Figure 3.4 shows the coordination of the phosphate anion. The phosphoryl oxygen atoms form hydrogen bonds to Oⁿ of Tyr32, N^ε of Arg42 of SGPB, and to water molecules W146 and W147. It should be noted that the observed phosphate binding site may merely be

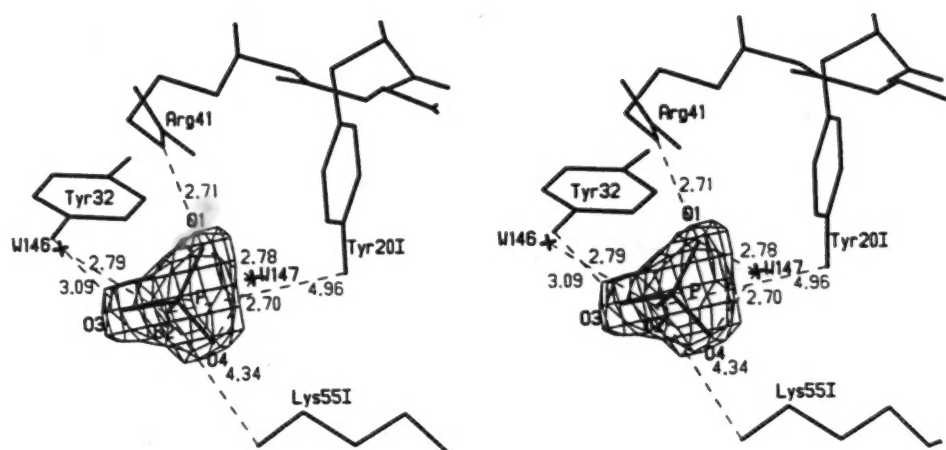


Figure 3.4 The phosphate binding site in the Leu¹⁸ variant complex, superimposed on a difference electron density map. The map contour level is 3σ . Dashed lines represent possible hydrogen bonds. Residues of OMTKY3 are labeled with a "I" following the sequence number to distinguish them from those of SGPB.

an artifact of the crystallization conditions (50 mM of phosphate buffer was used). Phosphate was not present in the measurement of the binding constants (Bigler, 1993; Lu, 1994). The presence of a partially occupied phosphate anion in the complex should not have a significant effect on the strength of the binding affinity, since it is relatively far from the reactive site. In addition, its presence in these structural studies should have little if any effect on the active site conformations.

Comparison of common water molecules in the complexes

In order to compare solvent structures in the S_1 pocket of different complexes, it is necessary to set up a distance criterion within which water molecules from different crystal structures are considered equivalent. For this purpose, crystallographically determined water molecules were grouped into three groups on the basis of their refined thermal factors (B-factors). The three B-factor shells include all water molecules, those with B-factors lower than $1.5 B_{\text{average}}$ and those with B-factors lower than B_{average} , respectively. Water molecules within the same B-factor shell were then compared in a pairwise manner. Table 3.2 gives the rms differences of the superpositions. The average B-factors are approximately $13.5 \text{ \AA}^2$ for main-chain atoms, $17.5 \text{ \AA}^2$ for side-chain atoms, and $37 \text{ \AA}^2$ for solvent molecules of the three complexes respectively.

To facilitate pairwise comparisons of water molecules, the numbering of water molecules in the three complexes was unified, following that of the previously reported structure (Read et al, 1983).

Table 3.2

The comparisons of solvent structures in the three complexes

	Average B factor ($\text{\AA}^2$)		Number of Water molecules		
	Mainchain	Sidechain	Water	Total	B<1.5 B _{Ave} B<B _{Ave}
Leu ¹⁸ variant complex	13.28	17.75	36.80	169	150 88
Ala ¹⁸ variant complex	13.71	17.68	37.47	182	174 86
Gly ¹⁸ variant complex	13.49	17.38	36.63	168	161 83
Number of common water molecules*				146	124 64
rms differences ($\text{\AA}$)					
Leu ¹⁸ versus Ala ¹⁸				0.38	0.30 0.24
Leu ¹⁸ versus Gly ¹⁸				0.34	0.29 0.20
Ala ¹⁸ versus Gly ¹⁸				0.41	0.30 0.25

* The numbering of water molecules in the three complexes follows that of the previously reported structure (Read et al., 1983), provided there is associated electron density for it at a 1.0 σ level in the refined $2|F_o| - |F_c|$ map and a refined B-factor lower than $60.0 \text{ \AA}^2$. If a water molecule is present in the three complexes with an identical ID, it is considered as a common water molecule.

A water molecule was assigned if there was associated density for it at a 1.0 σ level on the refined $2|F_o| - |F_c|$ map and if it had a refined B-factor lower than 60.0 $\text{\AA}^2$. If a water molecule was present with an identical ID in all the three complexes, it was considered as a common water molecule in the three complexes. There are 146 such common water molecules that are observed in all the three complexes, with the largest rms difference of 0.41 $\text{\AA}$ (Ala¹⁸ versus Gly¹⁸ variants, Table 3.2). Therefore a positional difference of 0.8 $\text{\AA}$ (2 times the rms difference) is significant enough and will be used as the distance cutoff in the following analysis of solvent structures. When comparing two structures, water molecules that superimposed within 0.8 $\text{\AA}$ are considered bound to an identical site and assigned an identical ID; differences greater than 0.8 $\text{\AA}$ are considered bound to different positions if different hydrogen-bonding patterns are also observed.

Redistribution of water molecules in the S₁ binding pocket

Comparisons among the three complexes and of each with the uncomplexed SGPB structure (Blanchard & James, 1995) reveal significant differences in the solvent distributions in the S₁ binding pocket. Figure 3.5 represents the superposition of the four structures, centering on the S₁ binding pocket. Water molecules bound in the S₁ binding pocket are summarized in Table 3.3. The numbering system for the solvent molecules of the uncomplexed SGPB follows that of Blanchard & James (1995), which is independent from that of the three complexes. In the S₁ binding pocket of the uncomplexed SGPB, there are five water molecules, B438, B440, B471, B529 and B543. These five water molecules

Table 3.3
Water molecules bound in the S₁ binding pocket of SGPB

Crystal Structure	Bound Water Molecules			
Uncomplexed SGPB	B440	B529	B438,	B543
B-factor ($\text{\AA}^2$)	41.91	64.75	47.50	55.17
Leu ¹⁸ variant complex	W42 (B440)	W71 (B529)	W15	W67
B-factor ($\text{\AA}^2$)	26.60	27.20	17.60	21.57
Ala ¹⁸ variant complex	W42 (B440)	W67,	W201	W202 (B438)
B-factor ($\text{\AA}^2$)	17.78	26.31	17.68	22.05
				37.65
				41.68
Gly ¹⁸ variant complex	W42 (B440)	W67,	W201	W202 (B438)
B-factor ($\text{\AA}^2$)	21.35	23.53	22.05	31.41
				53.55
				44.43

The numbers in parentheses denote equivalent water molecules in the uncomplexed SGPB (Blanchard & James, 1995).

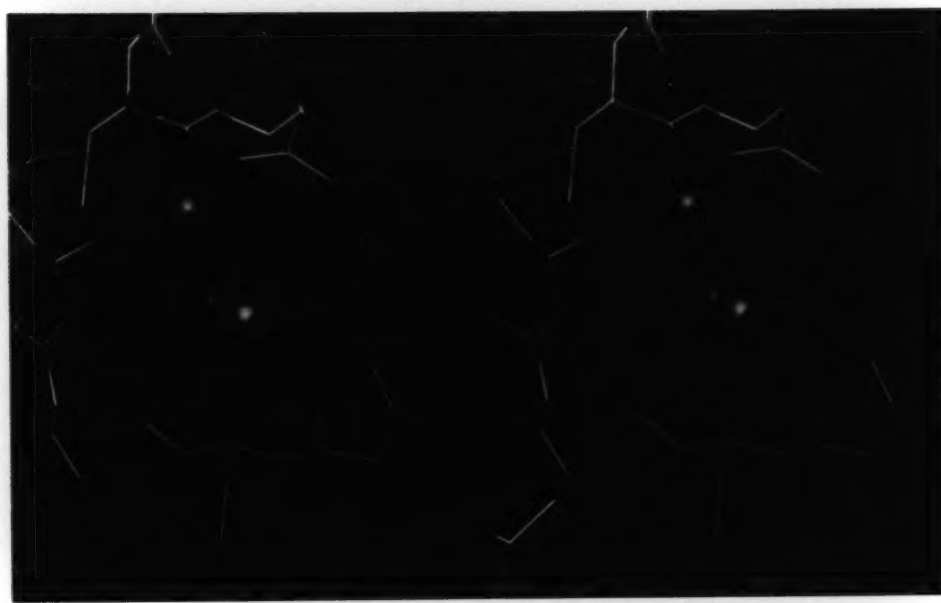


Figure 3.5 Redistribution of water molecules bound in the S_1 binding pocket. Superposition of four structures: the uncomplexed SGPB (in cyan), complexes of SGPB with the Leu¹⁸ variant (in green), the Ala¹⁸ variant (in magenta) and the Gly¹⁸ variant (in yellow). Residues of OMTKY3 are labeled with a "I" following the sequence number to distinguish them from those of SGPB. The picture is centered on the S_1 binding pocket to show the different positions of bound water molecules in each SGPB structure.

are conserved in the two non-crystallographically related molecules in the asymmetric unit of the uncomplexed SGPB (the rms difference for the five water molecules is 0.46 Å), therefore only one of the two SGPB molecules has been used in the structural comparisons and following discussion.

Upon binding of the Leu¹⁸ variant of OMTKY3, three water molecules (B438, B471 and B543) of the uncomplexed SGPB are excluded from the S₁ pocket by the side-chain of Leu18 of OMTKY3 (Fig. 3.5). The remaining two water molecules (B440 and B529) are actually equivalent to W42 and W71 of the Leu¹⁸ variant complex. In addition, two water molecules, W15 and W67 of the Leu¹⁸ variant complex, that were not observed in the uncomplexed SGPB, are present in the S₁ binding pocket. The complexes of the Ala¹⁸ and Gly¹⁸ variants display an almost identical solvent structure in the S₁ binding pocket (Fig. 3.5). Upon binding of the Ala¹⁸ (Gly¹⁸) variant to free SGPB, three water molecules (B471, B529 and B543) are excluded from the S₁ binding pocket. Water molecule B438, which was excluded by the Leu¹⁸ variant of OMTKY3, is in a position identical to W202 of the Ala¹⁸ and Gly¹⁸ variant complexes. Whereas water molecules W42 and W67 are conserved among the Leu¹⁸, Ala¹⁸ and Gly¹⁸ variant complexes, W15 and W71 are not observed in the Ala¹⁸ and Gly¹⁸ variant complexes. Three water molecules (W201, W203 and W204), none of which is observed in either the Leu¹⁸ variant complex, or the uncomplexed SGPB, are newly introduced to the Ala¹⁸ and Gly¹⁸ variant complexes. W42 (B440) is the only water molecule that is conserved in the uncomplexed SGPB and the three complexed SGPB molecules.

Water molecule B438 of the uncomplexed SGPB, which is also present in the Ala¹⁸ and Gly¹⁸ variant complexes (equivalent to W202), is excluded from the Leu¹⁸ variant complex (Fig. 3.5). This event may be coupled with the relatively small conformational change of loop Gly215 to Ser217 of SGPB in the Leu¹⁸ variant complex. Gly216 C of SGPB advances by 0.7 Å towards the side-chain of Leu18 of OMTKY3 in the complex. Consequently B438 is excluded from the S₁ binding pocket to avoid the otherwise too close contact (< 3.0 Å) with the C atom of Gly216. An alternative explanation might be that the exclusion of water molecule B438 by the side-chain of Leu18 of OMTKY3 results in a vacant space thereby triggering the inward shift of that loop of SGPB.

Water molecules participate in the enzyme-inhibitor association

In an attempt to appreciate the redistribution of water molecules in the S₁ binding pocket, we have examined the possible hydrogen bonds involving these water molecules. We found that all the water molecules that are newly introduced to the S₁ binding pocket are engaged in bridging hydrogen bonds between SGPB and OMTKY3. Table 3.4 lists all the possible hydrogen bonds observed in the interface of the complexes of SGPB and OMTKY3 variants, as well as those observed in the S₁ binding pocket of the uncomplexed SGPB. Figure 3.6 shows the possible hydrogen bonds between the Leu¹⁸ variant of OMTKY3 and SGPB, bridged by water molecules W15, W67 and W71. Asn36 N³² of OMTKY3 forms indirect hydrogen bonds with the carbonyl oxygen of Gly216 of SGPB, via water molecule W15. Asn36 O³¹ of OMTKY3 forms indirect hydrogen bonds with the carbonyl oxygen atoms of Ala192 and Gly218 of SGPB, via water molecules

Table 3.4

The hydrogen-bonding and non-bonded interactions (within 4.1Å) in the interface of the complexes of SGPB and OMTKY3 variants

H-bonding partners				H-bond distance (Å)		
				Leu ¹⁸	Ala ¹⁸	Gly ¹⁸
SGPB		OMTKY3				
Thr 39	O	Arg 21I	N ^{h2}	3.16	2.93	2.95
Arg 41	N	Tyr 20I	O	3.10	3.00	3.13
Arg 41	O	Tyr 20I	N	2.96	2.89	3.00
Tyr171	O	Lys 13I	N ²	2.88	3.19	3.18
Gly193	N	X 18I	O	2.50	2.61	2.47
Asp194	N	X 18I	O	3.3 ^a	3.23 ^a	3.29 ^a
Ser195	N	X 18I	O	3.05	2.94	3.12
Ser195	O ^o	X 18I	N	2.83	2.91	2.89
Gly216	N	Cys 16I	O	2.98	2.96	3.13
Gly216	O	Cys 16I	N	2.99	2.83	2.80
SGPB		Water				
Ala192	N	W42	O	2.82	2.89	2.85
Ala192	O	W42	O	3.12	2.80	2.84
Ala192	O	W71	O	2.59	N/A ^b	N/A ^b
Ala192	O	W202	O	N/A	3.71	3.54
Ala192	O	W203	O	N/A	3.23	3.65
Gly216	O	W15	O	2.99	N/A	N/A
Gly216	O	W204	O	N/A	3.1	3.04
Gly216	N	W202	O	N/A	3.48	3.49
Ser217	O	W42	O	2.82	3.02	2.99
Ser217	O	W202	O	N/A	3.57	3.87
Ser217	N	W202	O	N/A	3.38	3.56
Gly218	O	W67	O	3.08	3.03	2.92
Gly218	O	W71	O	2.74	N/A	N/A
Gly218	O	W42	O	3.99	3.28	3.43
Asn219	N	W42	O	4.03	3.29	3.41
Thr225	O	W202	O	N/A	3.08	3.42
Thr226	O ^{o1}	W42	O	3.45	2.81	3.96

Table 3.4/contd.

OMTKY3		Water				
Asn 36I	N ^{d2}	W15	O	3.17	N/A	N/A
Asn 36I	N ^{d2}	W201	O	N/A	2.93	3.07
Asn 36I	O ^{d1}	W67	O	2.86	2.79	3.08
Asn 39I	N	W26	O	3.40	3.40	3.55

Water		Water				
W15	O	W26	O	3.17	N/A	N/A
W15	O	W71	O	2.90	N/A	N/A
W26	O	W204	O	N/A	3.18	3.31
W42	O	W202	O	N/A	3.20	3.25
W67	O	W71	O	2.85	N/A	N/A
W201	O	W203	O	N/A	2.69	2.56
W202	O	W203	O	N/A	2.75	2.52
W204	O	W203	O	N/A	3.41	3.29

Uncomplexed	SGPB	Water	H-bond distance (Å)
Ala192	N	B440 (W42)	3.03
Ala192	O	B440 (W42)	3.20
Ala192	O	B529 (W71)	3.14
Glu192A	O	B471	3.21
Ser217	O	B440 (W42)	2.72
Gly218	O	B529 (W71)	2.61
Thr225	O	B438 (W202)	3.16
Thr226	O ^{d1}	B438 (W202)	2.91
B438 (W42)	O	B440 (W42)	2.70
B438 (W42)	O	B529 (W71)	3.36
B438 (W42)	O	B543	2.39
B543	O	B471	3.39
B543	O	B529 (W71)	2.98

* The Asp194 N of SGPB is in close contact with the X18I O of OMTKY3, but the value of the N-H...O angle precludes the formation of a hydrogen bond.

^b N/A refers to distances that are longer than 3.6 Å.

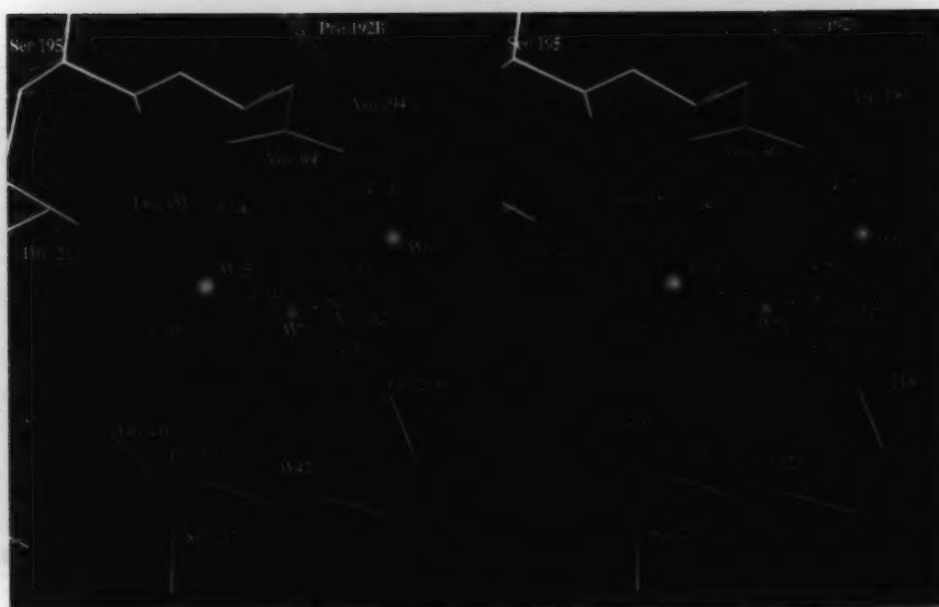


Figure 3.6 Water-mediated hydrogen bonds in the complex of SGPB with the Leu¹⁸ variant of OMTKY3. Dashed lines represent possible hydrogen bonds. Residues of OMTKY3 are labeled with a "I" following the sequence number to distinguish them from those of SGPB. The Leu¹⁸ variant complex (shown in yellow) is superimposed with the uncomplexed SGPB (shown in cyan), to emphasize that water molecules W15, W67 and W71, which are not present in the uncomplexed SGPB, are all engaged in bridging hydrogen bonds between SGPB and Asn36 of OMTKY3.

W67 and W71. Both W15 and W67 are newly introduced to the S_1 binding pocket upon the binding of the Leu¹⁸ variant. W71, which is equivalent to B529 of the uncomplexed SGPB, migrates by a distance of 0.5 Å, presumably to optimize these hydrogen-bonding interactions.

Figure 3.7 shows the probable hydrogen bonds observed in the Ala¹⁸ and Gly¹⁸ variant complexes, focusing only on those mediated by water molecules in S_1 binding pocket. The indirect hydrogen bonds between Asn36 of OMTKY3 and the carbonyl-oxygen atoms of Gly216, Ala192 and Gly218 of SGPB, as observed in the Leu¹⁸ variant complex, are also present in the Ala¹⁸ and Gly¹⁸ variant complexes. They occur via water molecules W201, W202, W203, W204 and W67. Note that water molecules W201, W203 and W204 are only present in the Ala¹⁸ and Gly¹⁸ variant complexes; there are no equivalent solvent sites in either the uncomplexed SGPB or the Leu¹⁸ variant complex. These water molecules are introduced to the S_1 binding pocket to fill the vacant space resulting from the small side-chains of Ala18 and Gly18 and perhaps more importantly, to bridge the hydrogen-bonding interactions between SGPB and OMTKY3.

*Correlation of binding free energy with the hydrophobic surface
area buried in the complex interface*

SGPB is a chymotrypsin-like serine proteinase with respect to its substrate specificity. The S_1 binding pocket in SGPB is hydrophobic in nature and relatively large in size. Therefore P_1 residues with bulky hydrophobic side-chains are preferred over those with small or polar side-chains. Non-coded amino acids, Abu, Ape, Ahx and Ahp were



Figure 3.7 Water-mediated hydrogen bonds in the complex of SGPB with the Ala¹⁸ variant of OMTKY3. Dashed lines represent possible hydrogen bonds. Residues of OMTKY3 are labeled with a "I" following the sequence number to distinguish them from those of SGPB. The Ala¹⁸ variant complex (shown in magenta) is superimposed with the uncomplexed SGPB (shown in cyan), to emphasize that water molecules W201, W203 and W204, which are not present in either the uncomplexed SGPB or the Leu¹⁸ variant complex, are all engaged in bridging hydrogen bonds between SGPB and Asn36 of OMTKY3.

introduced into the P_1 position of OMTKY3 to probe systematically the binding profile of the S_1 pocket. The negative binding free energy ($-\Delta\Delta G^\circ$) increases with the increasing number of side-chain atoms of the P_1 residue (Lu, 1994; Table 3.5). Due to the relatively low yields and high cost of chemical synthesis, the available quantity of the non-coded amino acid variants (Abu, Ape, Ahx and Ahp) was not sufficient for conducting crystallization trials. Therefore we have resorted to atomic models of the complexes of SGPB with these non-coded variants, in an effort to understand the S_1 specificity profile of SGPB. The torsion angles with which the P_1 side-chains were constructed are listed in Table 3.5. The P_1 side-chains of these variants can adopt favorable rotamers without causing any unacceptably close contacts with SGPB. The models were built on the basis of the Ala¹⁸ variant complex. The inward shift of residues Gly214 to Gly216, as observed in the Leu¹⁸ variant complex, does not cause any unfavorable contacts and has negligible effects on the subsequent calculation of accessible surface area. A few water molecules bound in the S_1 pocket, such as W15 and W71 in the Leu¹⁸ variant complex, or W201, W203 and W204 in the Ala¹⁸/Gly¹⁸ variant complex, are expected to be displaced by Ahp and possibly by Ahx.

The solvent accessible surface area was calculated for each complex, the free SGPB and the free OMTKY3 variants using the corresponding coordinates. The coordinates for the free SGPB and the free OMTKY3 variants were extracted from their complexes. The total solvent accessible surface area that is buried in the enzyme-inhibitor interface is defined as: $(S_E + S_I - S_{EI}) \text{ \AA}^2$, where S_E , S_I and S_{EI} are the

Table 3.5

The binding equilibrium constants (Lu, 1994) and calculated buried surface areas for a selection of complexes of SGPB and OMTKY3 variants

C	K_a (M^{-1})	$-\Delta G$ (Kcal/mol)	$-\Delta\Delta G$ (Kcal/mol)	Surface area ($\text{\AA}^2$)		Torsion angles (P_1 side-chain)			
				Apolar	Total	χ_1	χ_2	χ_3	χ_4
0	Gly	1.2×10^7	9.49	0	662 ^a	1165 ^a	-	-	-
1	Ala	3.6×10^6	11.47	1.98	719 ^a	1226 ^a	-	-	-
2	Abu	2.3×10^9	12.55	3.06	745 ^b	1253 ^b	-56°	-	-
3	Ape	1.2×10^{10}	13.51	4.02	772 ^b	1277 ^b	-56°	-50°	-
4	Ahx	2.1×10^{10}	13.84	4.35	784 ^b	1292 ^b	-56°	-50°	170°
5	Ahp	2.8×10^{10}	14.00	4.51	801 ^b	1306 ^b	-56°	-50°	170° 164°

a. Calculated from crystal structure.

b. Calculated from model structure.

total solvent accessible surface areas of the free SGPB, the free OMTKY3 and the SGPB-OMTKY3 complex, respectively.

Figure 3.8 is a plot of the binding free energy ($-\Delta\Delta G^\circ$, relative to that of the Gly¹⁸ variant) versus the buried hydrophobic surface area, for the six variants having straight aliphatic side-chains. The increase in $-\Delta\Delta G^\circ$ is directly correlated with the increase in the hydrophobic surface area buried in the complex interface. The resulting constant of proportionality (the slope of the fit) is $34 \text{ cal mol}^{-1} \text{ \AA}^{-2}$ for hydrophobic atoms (carbon and sulfur). Given the estimated error of $\pm 20\%$ for $-\Delta\Delta G^\circ$ (Lu, 1994), and the difference of $\pm 6\%$ in calculated solvent accessible surface areas between that of Connolly (Connolly, 1983) and that of Lee and Richards (Lee & Richards, 1971), the combined estimated error of $\pm 21\%$, ie, $\pm 7 \text{ cal mol}^{-1} \text{ \AA}^{-2}$ for the constant of proportionality.

As expected, similar correlations are also found for complexes of the same set of variants with α -chymotrypsin and SGPA (*Streptomyces griseus* proteinase A) and have comparable constants of proportionality ($37 \text{ cal mol}^{-1} \text{ \AA}^{-2}$ for α -chymotrypsin and $40 \text{ cal mol}^{-1} \text{ \AA}^{-2}$ for SGPA, data not shown). SGPA exhibits a substrate specificity very similar to that of SGPB. The observed strong correlations attribute the K_i variations to the amounts of the hydrophobic surface areas that are in contact in the enzyme-inhibitor interface. It indicates that the primary determinant of the binding free energy is the desolvation energy of the hydrophobic P₁ side-chain of the inhibitor and the hydrophobic S₁ binding pocket in these enzymes, when no conformational strains are involved.

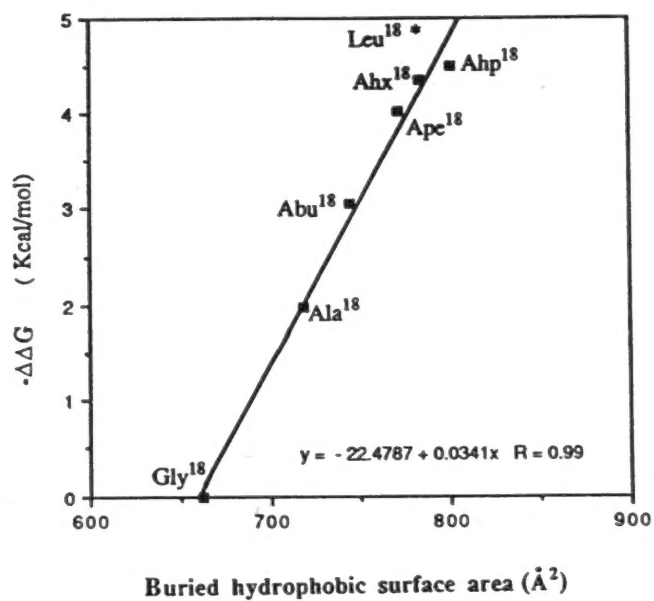


Figure 3.8 Linear correlation of the binding free energy $-\Delta\Delta G^\circ$ with the hydrophobic surface area buried in the interface of a selection of complexes of SGPB with OMTKY3 variants. The constant of proportionality is $34.1 \text{ cal mol}^{-1}\text{\AA}^{-2}$. The Leu¹⁸ variant (position shown by an "**") was not included in the line fitting.

DISCUSSION

The combination of the available thermodynamic data with the atomic structures showing the interactions of the P_1 variants of OMTKY3 with SGFB provides an excellent system to study the contribution of hydrophobic interactions to protein-protein associations. The currently-used measurements of the hydrophobic strength, by cavity-creating substitutions of interior residues, suffer from the complication of potential conformational disturbance of the protein structure (reviewed by Matthews, 1993; Fersht & Serrano, 1993). In contrast, the transfer of the P_1 residue to the preformed, relatively rigid S_1 pocket involves only displacement and redistribution of water molecules, thus providing a clean measurement of the unit contribution of hydrophobic interactions (hydrophobic strength) to protein and protein-complex stability.

There has been a large discrepancy (reviewed by Fersht & Serrano 1993; Matthews, 1993) over the value of the hydrophobic strength between the microscopic measurements (20 to 30 cal mol⁻¹ Å⁻² as derived from solute partition experiments, Chothia, 1974; Eisenberg & McLachlan, 1986) and macroscopic measurements (70 to 75 cal mol⁻¹ Å⁻² by surface tension measurement, Aveyard & Haydon, 1965). Substitutions of residues in the hydrophobic cores of proteins and subsequent assessments of the unfolding energy of the mutant proteins gave an estimate of 45 to 60 cal mol⁻¹ Å⁻² (Kellis et al., 1989; Shortle et al., 1990), which is close to the macroscopic value. Theoretical calculations applying size-difference correction (De Young & Dill, 1990; Sharp et al., 1991) and curvature-effects (Nicholls et al., 1991) tend to support the macroscopic

measurement. Our estimate (34 to 40 cal mol⁻¹ Å⁻²) is higher than the solvation parameters (20 to 30 cal mol⁻¹ Å⁻²) for hydrophobic atoms that were derived from experimental measurements of transferring free amino acids from water to a organic solvent (Chothia, 1975; Eisenberg & McLachlan, 1986), but lower than the hydrophobic strength (49 to 61 cal mol⁻¹ Å⁻²) estimated from substitutions of interior residues of barnase (Kellis et al, 1989). It would appear, therefore, that the S₁ pocket of SGPB is not as hydrophobic as the interior cavity in the core of a protein. This phenomenon can be understood by the unique chemical nature of the enzyme-inhibitor interface. Unlike the substitution-created interior cavities that are usually completely depleted of water molecules, the S₁ binding pocket is partially hydrated, as ordered water molecules are observed crystallographically. This partial hydration may account for the reduced hydrophobicity of the S₁ pocket. On the other hand, the apparently reduced hydrophobicity may also have to do with the fact that when an interior residue is substituted by a smaller amino acid, the surrounding residues tend to relax to fill the substitution-resulting cavity, a process in which the change of free energy may be considerable. Consequently, the substitution-resulting change of the unfolding energy is highly position-dependent. For example, the changes in the unfolding energies associated with the Leu to Ala substitutions in T4 lysozyme vary considerably from site to site, ranging from 1.9 Kcal/mol to 6 Kcal/mol (Eriksson et al, 1992). In the case of the SGPB-OMTKY3 complex, burial of the P₁ side-chain brings only displacement and/or redistribution of water molecules and minor, if any, conformational disturbances in the S₁ binding pocket. For this reason, the SGPB-OMTKY3 system should provide one of the most realistic

measurements of the hydrophobic strength. The constants of proportionality determined for SGPB, α -chymotrypsin and SGPA are comparable (34 to 40 cal mol⁻¹ Å⁻²), indicating that our assessment of the hydrophobic strength is reliable.

The homologous series of aliphatic, straight-chain P₁ variants, Gly¹⁸, Ala¹⁸, Abu¹⁸, Ape¹⁸, Ahx¹⁸ and Ahp¹⁸ provides a good system to estimate the contribution of a methyl group to protein-protein association and stabilization. The energetic contribution of a methyl group estimated from Gly to Ala substitution is higher than that estimated from Ala to Abu substitution (2.0 Kcal/mol versus 1.1 Kcal/mol). The discrepancy of 0.9 Kcal/mol may be due to the additional conformational flexibility of the P₁ glycine residue relative to those amino acids having side-chains. However, the observed correlations between the binding energy and the buried surface area suggest that the effect of the additional conformational variability of glycine is actually minimal. Otherwise, the Gly¹⁸ variant would be off the correlation curve by a considerable amount. The crystal structures of the Gly¹⁸ and the Ala¹⁸ variant complexes are virtually indistinguishable, glycine does not seem to behave differently from the other amino acids. It is mainly the relatively large increase in the buried hydrophobic surface area (an 8.6% increase for Gly to Ala substitution versus the 3.6% increase for Ala to Abu substitution) that accounts for the increase in binding energy with the addition of a methyl group to glycine.

The observed differences in the solvent structures for the different complexes lead to the conclusion that during the process of

SGPB-OMTKY3 complex formation bound water molecules in the S_1 binding pocket are not simply displaced by the inhibitor, but that there is an accompanying redistribution of solvent binding sites, synchronous with the conformational change of loop Gly215 to Ser217 of SGPB. We believe that these two events are to avoid unfavorable non-bonded contacts and perhaps more importantly, to compensate, to the maximum extent, for the entropic cost of ordering water molecules by introducing hydrogen bonds between SGPB and OMTKY3. Water-mediated hydrogen-bonding interactions are implicated in stabilizing a variety of other molecular complexes. Water-mediated hydrogen bonds have been observed in the recognition and formation of the DNA and trp repressor complex (Shakked et al., 1994), DNA and TATA binding protein complex (Kim & Burley, 1994). In these protein-DNA complexes bound water molecules become part of the recognition motif of the DNA sequence. In the case of protein and peptide association, it has been shown that water molecules play important roles in forming stable complexes of the OppA protein, an oligopeptide binding protein, with peptides of varying sequences and lengths (Tame et al., 1994). In the study of L-arabinose-binding protein, it was found that bound water molecules can modulate the specificity and affinity of the protein towards different sugar substrates (Quiocho et al., 1989). Additionally, Madden et al. (1992) have observed that a few conserved water molecules participate in forming key hydrogen bonds between the class I MHC molecule HLA-B27 and its bound peptide. Furthermore, Arnez and Steitz have proposed that pseudo-uridylation may stabilize tRNA^{Gln} through water-mediated linking of the pseudo-uridine to the backbone phosphate. The absence of these bound water molecules in the unmodified tRNA^(Gln) structure may account

for the reduced thermal stability of the unmodified tRNA^{Gln} (Arnez & Steitz, 1994). In our study, we have demonstrated that water molecules provide key interactions linking SGPB and the several OMTKY3 inhibitors. Taken together, the roles played by water molecules in protein-ligand recognition and association are extensive and should not be underestimated.

The water-mediated hydrogen bonds observed in the three complex structures show that Asn36 of OMTKY3 is invariably involved in forming water-bridged hydrogen bonds to residues of SGPB. Thus a mutation of Asn36 to the isosteric amino acid leucine could provide a measure of the impact on binding by the observed water-mediated hydrogen bonds. However, the energetic contributions of these bound water molecules are not expected to be high. Comparisons among avian ovomucoid inhibitor third domains from different species indicate that mutations Asn36 to Asp, Asn36 to Ser, and Asn36 to Ala do not result in much variation in the binding with SGPB (less than 0.9 Kcal/mol; Lu, 1994). The functions of these bound water molecules perhaps are to fill the vacant space left in the interface and to satisfy the hydrogen-bonding potentials of protein atoms. The gain of favorable enthalpic energy by forming hydrogen bonds may be comparable to the loss of unfavorable entropic energy by ordering the water molecules in the interface. The most desirable situation would be that the ligand displaces all the previously bound water molecules (entropically favorable), while concomitantly forming hydrogen-bonds or van der Waals contacts with the enzyme (enthalpically favorable).

Protein-protein associations usually involve large contact surface area (600 to 1800 Å²) and from 10 to 30 residues from either side. Mutagenesis studies in many systems have demonstrated that not all residues involved in the associations are energetically important. Usually, only a few of the residues are critical for tight binding. For example, two tryptophan residues constitute the "functional epitope" of the human growth hormone receptor, accounting for nearly three-quarters of the binding free energy (Clarkson and Wells, 1995). The functional epitopes of the hormone receptor are analogous to the specificity determinant, the P₁ residue of OMTKY3, in terms of energy contributions as well as their geometrical disposition relative to the main-body of the protein (Trp104 and Trp169 project from two adjacent loops). Large hydrophobic amino acids are increasingly implicated in protein-protein association. The crystal structure of N cadherin suggests that a tryptophan residue may play an important role in cadherin dimerization, a process thought to enhance cell-adhesion (Shapiro et al., 1995). Accordingly, large surface depressions that accommodate those bulky hydrophobic residues become interesting when improved specificity and/or enhanced affinity are concerned. Such surface depressions bear a strong physical resemblance to the S₁ binding pocket of SGPB. Since functional epitopes are the primary targets for inhibitor design, knowledge obtained from OMTKY3 may be applicable to designing ligands targeting hormone receptor proteins, MHC molecules and other functionally important proteins. We have demonstrated that the P₁ substitution-associated changes in buried hydrophobic surface area are the direct cause of the variations in the binding energy of OMTKY3 with SGPB, α-chymotrypsin and SGPA. It is reasonable to believe that the buried

hydrophobic surface area-binding free energy correlation is general to other molecular complexes in which large hydrophobic depressions are involved. It would be very interesting to see what values the constant of proportionality would be in other systems, to identify the factors that influence the hydrophobic strength. It will help to resolve the controversy over the values of the unit contribution of hydrophobic effect and further our understanding about the nature of hydrophobic effect to protein stability.

MATERIALS AND METHODS

Crystallization

SGPB, purified according to the procedure of Jurasek et al. (1979), was kindly provided by Dr. L. B. Smillie. The cloning, overexpression system and the purification procedures for OMTKY3 have been described (Lu et al, 1992; Lu, 1994). In addition to the P₁ substitutions, the recombinant variants are five-amino acids shorter at the N-terminus than native OMTKY3. The 1:1 stoichiometric complexes of SGPB-OMTKY3 variants were prepared by mixing SGPB with each OMTKY3 variant in a 1:1.05 molar ratio of enzyme to inhibitor. Crystals of the complex were grown by hanging drop methods, under conditions similar to those previously reported (Fujinaga et al, 1982). Each drop contained 3 μ l of protein solution (10mg/ml) and an equal volume of reservoir solution (2% of PEG 6000, 50 mM sodium/potassium phosphate at pH 6.5). Seeding was usually required to get good crystals. The crystals reach their final size (1.0 x 0.3 x 0.1 mm³) in about a week. These crystals are isomorphous to

those of the complex of SGPB with native OMTKY3. Unit cell parameters of the crystals of the three complexes (Leu¹⁸, Ala¹⁸ and Gly¹⁸) are listed in Table 3.6.

Data collection

X-ray diffraction data for the Leu¹⁸ variant and the Gly¹⁸ variant complexes were collected on San Diego MultiWire System twin area detectors. Graphite monochromated Cu K_α radiation was generated with a Rigaku rotating anode generator RU-200BH operating at 40 kV and 150mA. The 1.8 Å data set was collected from a single crystal of the Leu¹⁸ variant complex, the 1.9 Å data set was collected from a single crystal of the Gly¹⁸ variant complex. The data collections were carried out at 24°C, each lasting about 2 days. The 1.4 Å data set of the Ala¹⁸ variant complex was collected at beam-line 18B of the Photon Factory in Japan (room temperature 18°C), using Weissenberg geometry with a crystal-to-image-plate distance of 286.50 mm (Sakabe, 1983). The recorded reflections were indexed and integrated by profile fitting with the program WEIS (Higashi, 1989). The data-collection statistics for all these complexes are given in Table 3.6.

Structure determination and refinement

The phases for the complexes of SGPB with the three OMTKY3 variants were taken from those of the complex of SGPB with native OMTKY3 (Read et al., 1983). A $(2|F_o| - |F_c|)\exp(i\alpha_{c1})$ map, where $|F_o|$ is the observed structure-factor amplitudes of the variant complex, $|F_c|$ and α_{c1} are the structure-factor amplitudes and phases calculated from the structure

Table 3.6.

Statistics of data collection

Space group	P2 ₁			
Cell dimensions	a (Å)	b (Å)	c (Å)	β (°)
Leu ¹⁸ complex	45.53	54.66	45.59	119.19
Ala ¹⁸ complex	45.38	54.62	45.47	119.20
Gly ¹⁸ complex	45.48	54.68	45.55	118.99
Data collection	Leu ¹⁸	Ala ¹⁸	Gly ¹⁸	
Maxi. resolution (Å)	1.8	1.4	1.9	
No. of crystals used	1	1	1	
Total No. of observations	41,124	83,250	36,307	
No. of unique reflections	17,183	30,783	13,125	
Average redundancy	2.4	2.7	2.8	
Completeness of data				
overall (%)	94.3	80.3	84.9	
Highest resolution shell(%)	51.1	56.0	36.5	
	(1.9-1.8Å)	(1.5-1.4Å)	(2.0-1.9Å)	
R _{merge} (%)	5.29	7.33	4.29	

$$R_{\text{merge}} = \frac{\sum_{hkl} [(\sum_i I_i - \langle I \rangle) \sum_i I_i]}{\sum_{hkl} \sum_i I_i}$$

of Read, et al. (1983), was calculated with the program SIGMAA (Read, 1986) and displayed with FRODO (Jones, 1986). The P₁ residue Leu18 was replaced by the appropriate P₁ amino acid side-chains to fit best the observed electron density. The starting R-factor for the Leu¹⁸, Ala¹⁸ and Gly¹⁸ variant complexes were 36.4%, 42.5% and 35.9%, respectively, and were improved to 20.6%, 22.8% and 20.8% respectively after energy minimization and simulated annealing with XPLOR (Brunger, 1986). The model was then subjected to TNT refinement (Tronrud, 1992). All reflections with $d_{\min} < 20.0 \text{ \AA}$ were used for the refinements. After ten cycles of TNT refinement, a new $2|F_o| - |F_c|$ map was made for visual inspection. Special attention was given to regions where there were noticeable conformational disparities with the previously reported structure and where atoms displayed high thermal factors. Up to this stage no water molecules were built into the complex model. A $(|F_o| - |F_c|)\exp(i\alpha_{\text{calc}})$ difference map was made to assign water molecules. Water molecules that were present in the previously reported structure were built into the variant complex model if there was associated densities for it above the 2.5σ contour level in the difference map. The manually-adjusted model was subjected to another 10 cycles of TNT refinement. Afterwards water molecules with B factors higher than $50 \text{ \AA}^2$ were examined closely. A water molecule would be deleted if there was no associated density at the 1σ contour level on a $2|F_o| - |F_c|$ map and if the B factor was greater than $60 \text{ \AA}^2$. New water molecules were added if there was electron density above the 2.5σ level on a $|F_o| - |F_c|$ difference map. The geometry of potential hydrogen bond donor or acceptor atoms was also checked to assure that it was meaningful to introduce a water molecule in that region. Refinement was judged to

reach convergence typically after a total of 30 cycles of TNT refinement.

Structure comparison

Superpositions of the several structures were done in a pairwise manner using the lsq commands in the program "O" (Jones et al, 1991). The root-mean-square (rms) differences for common atoms were obtained subsequently from the program. Crystallographically determined water molecules were grouped into three shells on the basis of their thermal factors. The three B-factor shells include all water molecules, water molecules with B-factor lower than the B_{average} and lower than $1.5 B_{\text{average}}$, respectively. Water molecules within the same B-factor shell are compared.

Modeling of the non-coded amino acid variants of OMTKY3

The currently available quantities of the non-coded amino acid variants (Abu, Ape, Ahx and Ahp) were not sufficient for conducting crystallization trials. Accordingly, we have built models of the four complexes on the basis of the crystal structure of the complex between SGPB and the Ala¹⁸ OMTKY3 variant. In doing so, the side-chains of the P₁ residues were rotated to most favorable rotamer conformations while maintaining favorable van der Waals contacts. The modeling was very straight forward, no energy minimization was required or applied.

Calculation of solvent accessible surface area

Solvent accessible surface area was calculated using the program MS (Connolly, 1983). The probe radius for water molecules was 1.4 Å. Van

der Waals radii for protein atoms are the default of the program. Sulfur is treated as hydrophobic. Solvent accessible surface areas for the free SGPB and the free OMTKY3 variants were calculated with the coordinates extracted from the corresponding complexes, under the valid assumption that the binding process is close to a rigid-body association.



REFERENCES

- Ardelt WA & Laskowski M Jr. 1991. Effect of single amino acid replacements on the thermodynamics of the reactive site peptide bond hydrolysis in ovomucoid third domain. *J Mol Biol* 220:1041-1053.
- Arnez JG & Steitz TA. 1994. Crystal structure of unmodified tRNA^{Gln} complexed with glutamyl-tRNA synthetase and ATP suggests a possible role for pseudo-uridines in stabilization of RNA structure. *Biochemistry* 33:7560-7567.
- Aveyard R & Haydon DA. 1965. Thermodynamic properties of aliphatic hydrocarbon/water interfaces. *Trans Faraday Soc* 61:2255-2261.
- Bigler TL, Lu W, Park SJ, Tashiro M, Wieczorek WR & Laskowski M Jr. 1993. Binding of amino acid side chains to preformed cavities: Interaction of serine proteinases with turkey ovomucoid third domains with coded and noncoded P₁ residues. *Protein Science* 2:786-799.
- Blanchard H & James MNG. 1995. X-ray crystallographic structure of *Streptomyces griseus* proteinase B at its enzymatically optimal pH. Manuscript in preparation.
- Bode W, Epp O, Huber R, Laskowski M Jr & Ardelt W. 1985. The crystal and molecular structure of the third domain of silver pheasant ovomucoid (OMSVP3). *Eur J Biochem* 147:387-395.
- Bode W, Wei AZ, Huber R, Meyer E, Travis J & Neumann S. 1986. X-ray crystal structure of the complex of human leukocyte elastase (PMN elastase) and the third domain of the turkey ovomucoid inhibitor. *EMBO J* 5:2453-2458.
- Bode W & Huber R. 1992. Natural protein proteinase inhibitors and their interactions with proteinases. *Eur J Biochem* 204:433-451.

- Clackson T & Wells JA. 1995. A hot spot of binding energy in a hormone-receptor interface. *Science* 267:383-386.
- Chothia C. 1974. Hydrophobic bonding and accessible surface area in proteins. *Nature* 248:338-339.
- Codding PW, Delaere LT, Hayakawa K, Hutcheon WL, James MNG & Jurasek L. 1974. The 4.5 Angstrom resolution structure of a bacterial serine proteinase from *Streptomyces griseus*. *Can J Biochem* 52:208.
- Connolly ML. 1983. Analytical molecular surface calculation. *J Appl Cryst* 16:548-558.
- De Young LR & Dill KA. 1990. Partitioning of nonpolar solutes into bilayers and amorphous n-alkanes. *J Phys Chem* 94:801-809.
- Eisenberg D & McLachlan AD. 1986. Solvation energy in protein folding and stability. *Nature* 319:199-203.
- Eriksson AE, Baase WA, Zhang XJ, Heinz DW, Blaber M, Baldwin EP & Matthews BW. 1992. Response of a protein structure to cavity-creating mutations and its relation to the hydrophobic effect. *Science* 255:178-183.
- Fersht NR & Serrano L. 1993. Principles of protein stability derived from protein engineering experiments. *Curr Opin Struc Biol* 3:75-83.
- Fujinaga M, Sielecki AR, Read RJ, Ardelt W, Laskowski M Jr & James MNG. 1987. Crystal and molecular structures of the complex of α -chymotrypsin with its inhibitor turkey ovomucoid third domain at 1.8 Å resolution. *J Mol Biol* 195:397-418.
- Higashi T. 1989. The processing of diffraction data taken on a screenless Weissenberg camera for macromolecular crystallography. *J Appl Crystallogr* 22:9-18
- Jones TA. 1985. Interactive computer graphics: FRODO. *Meth Enzym* 115:157-171. With modifications for TOM UQV/CIT v.2.8.0 by Camillau C, Israel M, Griffin G & Chirino A.

- Jones TA, Zou JY, Cowan SW, Kjeldgaard M. 1991. Improved methods for building protein models in electron density maps and the location of errors in these models. *Acta Cryst A* 47:110-119.
- Jurasek L, Johnson P, Olafson RW & Smillie LB. 1971. An improved fractionation system for pronase on CM-sephadex. *Can J Biochem* 49:1195-1201.
- Kabsch W & Sander C. 1983. Dictionary of protein secondary structure: pattern recognition of hydrogen bonded and geometrical features. *Biopolymers* 22:2577-2637.
- Kellis JT Jr, Nyberg K & Fersht AR. 1989. Energetics of complementary side-chain packing in a protein hydrophobic core. *Biochemistry* 28:4914-4922.
- Kim JL & Burley SK. 1994. 1.9 Å resolution refined structure of TBP recognizing the minor groove of TATAAAG. *Nature Struct Biol* 1:638-651.
- Kollman PA. 1994. Theory of macromolecule-ligand interactions. *Curr Opin Struct Biol* 4:240-245.
- Komiyama T, Bigler TL, Yoshida N, Noda K & Laskowski M Jr. 1991. Replacement of P₁ Leu¹⁸ by Glu¹⁸ in the reactive-site loop of turkey ovomucoid third domain converts it into a strong inhibitor of Glu-specific *Streptomyces griseus* proteinase (GluSGP). *J Biol Chem* 266:10727-10730.
- Laskowski M Jr & Kato I. 1980. Protein inhibitors of proteinases. *Biochemistry* 49:593-626.
- Lee B & Richards FM. 1971. Interpretation of protein structures: estimation of static accessibility. *J. Mol. Biol.* 55:379-400.
- Lu W, Zhang W, Molly SS, Thomas G, Ryan K, Chiang Y, Anderson S & Laskowski M Jr. 1993. Arg¹⁵-Lys¹⁷-Arg¹⁸ turkey ovomucoid third domain inhibits human furin. *J Biol Chem* 268:14583-14585.

- Lu W. 1994. Energetics of the interactions of ovomucoid third domain variants with different serine proteinases. *Ph.D. Thesis*, Purdue University, West Lafayette, Indiana.
- Madden DR, Gorga JC, Strominger JL & Wiley DC. 1992. The three-dimensional structure of HLA-B27 at 2.1 Å resolution suggests a general mechanism for tight binding to MHC. *Cell* 70:1035-1048.
- Matthews BW. 1993. Structural and genetic analysis of protein folding and stability. *Curr Opin Struc Biol* 3:589-593.
- Musil D, Bode W, Huber R, Laskowski M Jr, Lin TY & Ardelt W. 1991. Refined X-ray crystal structures of the reactive site modified ovomucoid inhibitor third domains from silver pheasant (OMSVP3*) and from Japanese Quail (OMJPQ3*). *J Mol Biol* 220:739-755.
- Nicholls A, Sharp KA & Honig B. 1991. Protein folding and association: insights from the interfacial and thermodynamic properties of hydrocarbons. *Proteins* 11:282-296.
- Papamokos E, Weber E, Bode W, Empie MW, Kato I & Laskowski M Jr. 1982. Crystallographic refinement of Japanese quail ovomucoid, a-Kazal type inhibitor, and model building studies of complexes with serine proteases. *J Mol Biol* 158:515- 537.
- Quirocho FA, Wilson DK & Vyas NK. 1989. Substrate specificity and affinity of a protein modulated by bound water molecules. *Nature* 340:404-406.
- Read RJ, Fujinaga M, Sielecki AR & James MNG. 1983. Structure of the complex of *Streptomyces griseus* protease B and the third domain of the turkey ovomucoid inhibitor at 1.8 Å resolution. *Biochemistry* 22:4420-4433.
- Read RJ & James MNG. 1986. Introduction to the protein inhibitors: X-ray crystallography. In *Proteinase inhibitors* (Barrett AJ and Salvesen G, Eds.) pp. 301-336. Elsevier, Amsterdam.

- Read RJ. 1986. Improved Fourier coefficients for maps using phases from partial structures with errors. *Acta Cryst A* 42:140-149.
- Schechter I & Berger A. 1967. On the size of the active site in proteases. I. Papain. *Biochem Biophys Res Commun* 27:157-162.
- Shakked Z, Guzikovich-Guerstein G, Frolow F, Rabinovich D, Joachimiak A & Sigler PB. 1994. Determinants of repressor/operator recognition from the structure of the trp operation binding site. *Nature* 368:469-473.
- Shapiro L, Fannon AM, Kwong PD, Thompson A, Lehmann MS, Grubel G, Legrand J, Als-nielsen J, Colman DR & Hendrickson WA. 1995. Structural basis of cell-cell adhesion by cadherins. *Nature* 374:327-336.
- Sharp KA, Nicholls A, Friedman R & Honig B. 1991. Extracting hydrophobic free energies from experimental data: relationship to protein folding and theoretical models. *Biochemistry* 30:9686-9697.
- Shortle D, Stites WE & Meeker AK. 1990. Contributions of the large hydrophobic amino acids to the stability of Staphylococcal nuclease. *Biochemistry* 29:8033-8041.
- Tame JRH, Murshudov GN, Dodson EJ, Neil TK, Dodson GG, Higgins CF & Wilkinson AJ. 1994. The structural basis of sequence-independent peptide binding by OppA protein. *Science* 264:1578-1581.
- Tronrud DE. 1992. Conjugate-direction minimization: An improved method for the refinement of macromolecules. *Acta Cryst A* 48:912-916.
- Weber E, Papamokos E, Bode W, Huber R, Kato I. & Laskowski M Jr. 1981. Crystallization, crystal structure analysis and molecular model of the third domain of Japanese Quail ovomucoid, a Kazal type inhibitor. *J Mol Biol* 149:109-123.
- Wieczorek M & Laskowski M Jr. 1983. Formation of covalent hybrids from amino- terminal and carboxy-terminal fragments of two ovomucoid third domains. *Biochemistry* 22:2630-2636.

Wieczorek M, Park SJ & Laskowski M Jr. 1987. Covalent hybrids of ovomucoid third domains made from one synthetic and one natural peptide chain. *Biochem Biophys Res Commun* 144:499-504.

Chapter 4

Estimating the Energetic Effects of Burying a Salt bridge in a Hydrophobic Pocket: Crystal Structures of the Glu¹⁸, Gln¹⁸, Asp¹⁸ and Asn¹⁸ Variants of Turkey Ovomuroid Third Domain in Complex with *Streptomyces griseus* Proteinase B at Various pHs

INTRODUCTION

The "standard mechanism" whereby canonical protein inhibitors of the serine proteinases inactivate their cognate enzymes by forming a very tight complex (Laskowski & Kato, 1980) is firstly an issue of macromolecular recognition. A canonical protein inhibitor is mistaken for a substrate by its cognate enzyme and is bound uncleaved in a non-covalent association. The reactive-site peptide bond is positioned, with the assistance of the inhibitor scaffolding, at a distance that is too far away from Ser O^γ for nucleophilic attack to take place (Reviewed by Read & James, 1986; Bode & Huber, 1992). Turkey ovomucoid inhibitor third domain (OMTKY3) is one of the best studied canonical inhibitors of the serine proteinases. How OMTKY3 interacts with its cognate enzymes, such as *Streptomyces griseus* Proteinase B (SGPB) in our study, will provide important insights for understanding protein-protein

associations in general. The binding process of OMTKY3 to SGPB is basically a rigid-body association, involving only small conformational changes in either enzyme or inhibitor. Thus it represents a simplified model with which to relate the binding free energy to structural data. The binding constants have been obtained by measuring the residual proteolytic activity of SGPB following inhibition by the OMTKY3 variants (Bigler, et al., 1993; Lu, 1994). The interpretation of the thermodynamic data will be greatly assisted by the high resolution crystal structures of the various OMTKY3 variants in molecular complexes with SGPB (Huang et al, 1995), native OMTKY3 in complexes with SGPB (Read et al, 1983), α -chymotrypsin (Fujinaga et al, 1987) and human leukocyte elastase (Bode et al., 1986).

The ovomucoids from avian eggs are multidomain proteins, typically with three domains each one capable of inhibiting a serine proteinase. OMTKY3 is the third domain of the ovomucoid from turkey having 56 amino acids and Leu18I in the P_1 position. It strongly inhibits serine proteinases with a hydrophobic S_1 pocket, such as chymotrypsin, subtilisin and elastase. The exposed P_1 Leu18I of OMTKY3 inserts into the preformed S_1 pocket of the enzyme upon complex formation (Read et al., 1983; Fujinaga et al., 1987). In an effort to establish a sequence-reactivity algorithm for OMTKY3, a series of recombinant variants of this inhibitor have been generated, in which 20 coded and 5 non-coded amino acids were introduced to the P_1 position. The association constants of these 25 P_1 variants of OMTKY3 have been measured for SGPB and for five other serine proteinases (Bigler et al, 1993; Lu, 1994). SGPB is a chymotrypsin-like proteinase with regards to the three-dimensional structure (James et al, 1978) and the substrate

specificity (Jurasek et al., 1976). Naturally-occurring OMTKY3 (with Leu18I at the P₁ position) forms an extremely tight complex with SGPB, with an association constant (K_a) of $5.6 \times 10^{10} \text{ M}^{-1}$. This is equivalent to 14.4 Kcal/mol of binding free energy (Bigler et al, 1993; Lu, 1994). The molecular scaffolding of OMTKY3 provides about 9.5 Kcal/mol of binding free energy, as displayed by the Gly¹⁸ variant which has no P₁ side-chain (Lu, 1994; Huang et al., 1995). The favorable hydrophobic interactions between the P₁ leucine side-chain and the S₁ pocket thus provide an additional 4.9 Kcal/mol stability to the complex. In contrast, polar or charged P₁ residues destabilize the complex. The substitution of P₁ Leu by the isosteric amino acid Asn destabilizes the complex by 3.3 Kcal/mol; the substitution of P₁ Asn by the ionizable side-chain of P₁ Asp further destabilizes the complex by an additional 2.2 Kcal/mol.

X-ray crystallographic studies are underway in our laboratory to determine and analyze the structures of the complexes between SGPB and the 20 coded P₁ variants of OMTKY3. These studies are to understand the exquisite discriminating ability of OMTKY3 by examining enzyme-inhibitor interactions. To address the contribution of hydrophobic interactions to the free energy of binding, we have reported the structures of the Leu¹⁸, Ala¹⁸ and Gly¹⁸ variants of OMTKY3 in complex with SGPB (Huang et al, 1995). Here we report the structures of the Glu¹⁸, Gln¹⁸, Asp¹⁸ and Asn¹⁸ variants in complex with SGPB, in an attempt to understand the role of electrostatic interactions in enzyme-inhibitor recognition and association processes. The association equilibrium constants (measured at pH 8.3) for these complexes are $2.3 \times 10^6 \text{ M}^{-1}$ (Glu¹⁸ variant), $6.9 \times$

10^8 M^{-1} (Gln¹⁸ variant), $3.9 \times 10^6 \text{ M}^{-1}$ (Asp¹⁸ variant) and $1.8 \times 10^8 \text{ M}^{-1}$ (Asn¹⁸ variant).

It has long been recognized that ionizable residues often exist in the uncharged form when buried in a hydrophobic environment, thus displaying a large shift in pK_a from their solvated state (Rupley et al., 1967; and recently reviewed by Yang & Honig, 1992). Acquisition of charge usually triggers an unfolding of the protein or a dissociation of a molecular complex. As shown in Figure 4.1, the S_1 binding pocket of SGPB is highly hydrophobic. Understandably, Glu18I of the Glu¹⁸ variant of ONTKY3 exhibits a huge increase in pK_a of 4.2 pH units, from pK_a of 4.5 in the free inhibitor, to pK_a of 8.7 when complexed with SGPB (Qasim et al, 1995). Asp18I of the Asn¹⁸ variant exhibits a even larger pK_a shift (5.1 pH units), from pK_a of 4.3 in the free inhibitor, to pK_a of 9.4 when complexed with SGPB (Qasim & Laskowski, unpublished data). In their experiments, the association constants of SGPB interacting with the Glu¹⁸, Asp¹⁸ and Gln¹⁸ variants were measured at various pH points, ranging from pH 3.0 to 10.5 with an interval of ~0.5 pH unit. The K_a s of the Glu¹⁸ and Asp¹⁸ variants were calibrated with those of the non-ionizable Gln¹⁸ variant, and the pK_a of Glu18I and Asp18I were derived from the pH dependent curve of the association constants (Qasim et al, 1995). When the Glu18I carboxyl group is protonated both in the free and in the complexed forms, the K_a for the Glu¹⁸ variant interacting with SGPB is about $1 \times 10^{10} \text{ M}^{-1}$, as opposed to the K_a of $6 \times 10^5 \text{ M}^{-1}$ when the Glu18I carboxyl group carries a negative charge both in free and in complexed forms (Qasim et al., 1995). It is very surprising that the complexes between SGPB and the Glu¹⁸ and Asp¹⁸ variants remain strongly

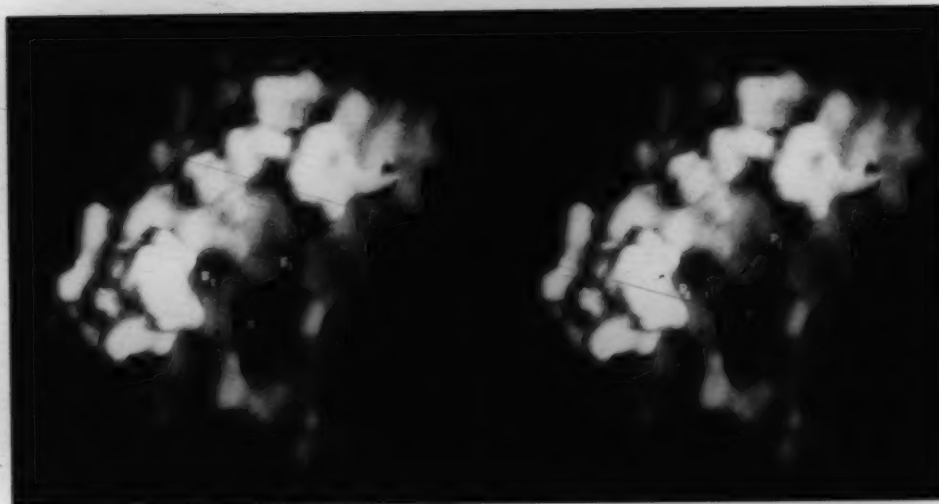


Figure 4.1 (a). The electrostatic potentials on the van der Waals surface of SGPB. The negative electrostatic potentials on the molecular surface are shown in red, and the positive potentials shown in blue. Hydrophobic surface areas are shown in white. The S_2 , S_1 and S_1' binding sites are labeled. The S_1 binding pocket of SGPB is predominantly hydrophobic.

(b). The insertion of the P_1 side-chain into the S_1 binding pocket. Only the P_2 (Thr17I), P_1 (Glu18I) and P_1' (Glu19I) residues of the Glu¹⁸ variant of OMTKY3 are shown.

associated with an negatively-charged carboxylate buried in the highly hydrophobic S_1 pocket.

To visualize the complexes between SGPB and OMTKY3 when the P_1 Glu18I and Asp18I side-chains are in the neutral and the charged states, and to understand how a negatively-charged P_1 side-chain can be accommodated in the hydrophobic S_1 pocket, we have determined the structures of the complexes at pH 6.5 and pH 10.7, which are 2.2 pH unit below and 2 pH unit above the pK_a of Glu18I of OMTKY3. The crystals of these complexes were grown initially at pH 6.5, a value at which the side-chains of Glu18I and Asp18I are predominantly protonated (neutral). The crystals were then soaked with mother liquor buffered at pH 10.7, a value at which the carboxyl groups of Glu18I and Asp18I are essentially completely deprotonated and carry a negative charge. As a control experiment, the structure of the complexes between SGPB and the Gln¹⁸ variant, the non-ionizable counterpart of the Glu¹⁸ variant, was also determined at pH 6.5 and 10.7. The structure of the complex of Asn¹⁸ variant with SGPB was determined only at pH 6.5.

Interestingly, a potassium ion has found in the pH 10.7 structures of the Glu¹⁸ and Asp¹⁸ variants, but not in any of the other structures. The K^+ ion coordinates with the negatively-charged carboxylate group of the P_1 side-chain in both structures. Given the observed unusually high pK_a s of Glu18I and Asp18I in the complexed state, the presence of a cation, K^+ , in the pH 10.7 structures and the absence of it in the pH 6.5 structures of the Glu¹⁸ and Asp¹⁸ variants is a clear indication that the side-chains of P_1 Glu18I and Asp18I are neutral at pH 6.5, and are ionized (negatively-charged) at pH 10.7. During the deprotonation

event, the P₁ side-chain acquires a negative charge and attracts a counter ion, K⁺ in this case, to form a salt bridge, thereby preventing the dissociation of the complexes. The energy cost of burying a salt bridge in a highly hydrophobic environment can be estimated by measuring the stability of the SGPB-OMTKY3 complexes at various pHs.

In this paper, structures that were determined at pH 6.5 (neutral pH) will be labeled with an "(N)", whereas those determined at pH 10.7 (high pH) will be labeled with an "(H)", to distinguish them from one another. Residues of OMTKY3 are labeled with a "I" following their sequential residue number, to distinguish them from the residues of SGPB.

RESULTS

Quality of the structures of various complexes

The crystal structures of the complexes between SGPB and the Glu¹⁸, Gln¹⁸ and Asp¹⁸ variants of OMTKY3 have been determined at pH 6.5 and pH 10.7. The crystal structure of the Asn¹⁸ variant has been determined at pH 6.5. These structures were refined with 1.8 Å resolution diffraction data. The crystallographic R-factors, the final stereochemical parameters and the average thermal factors of the current refined molecular structures are listed in Table 4.1. The final R-factors for the pH 6.5 structures of the complexes of the Glu¹⁸, Gln¹⁸, Asp¹⁸ and Asn¹⁸ variants are all comparable, at 15.1% , 14.9%, 15.7% and 15.9%, respectively. These are values expected for well refined, medium to high resolution crystal structures of proteins. On the other hand,

Table 4.1. Statistics of structural refinement

	Glu ¹⁶ (N)	Glu ¹⁶ (H)	Gln ¹⁶ (N)	Gln ¹⁶ (H)	Asp ¹⁶ (N)	Asp ¹⁶ (H)	Asn ¹⁶
No. of Reflections used	16,979	16,886	17,441	17,120	17,035	17,130	17,249
Resolution range (Å)	20-1.8	20-1.8	20-1.8	20-1.8	20-1.8	20-1.8	20-1.8
Completeness (%)	91	90	93	92	91	92	92
R-factor* (%)	15.1	19.8	14.9	17.5	15.7	19.6	15.9
rms deviations from ideal stereochemistry							
bond distance (Å)	0.022	0.024	0.022	0.023	0.024	0.025	0.024
bond angle (°)	2.5	2.6	2.5	2.6	2.6	2.8	2.6
planar groups (Å)	0.021	0.020	0.021	0.020	0.021	0.021	0.021
No. of protein atoms	1698	1698	1698	1698	1697	1697	1697
No. of water atoms	171	151	156	153	170	151	161
No. of phosphate anions	1	1	1	1	1	1	1
Average B-factors (Å ²)							
main-chain	11.2	12.5	11.4	10.2	11.7	12.0	12.8
side-chain	15.5	16.2	15.5	14.3	15.8	14.9	17.5
solvent	34.6	34.0	32.9	31.0	33.3	31.0	35.9

$$^*R\text{-factor} = \sum_{hkl} | |F_o| - |F_c| | / \sum_{hkl} |F_o|$$

the current R-factors for the pH 10.7 structures of the complexes of the Glu¹⁸ and Asp¹⁸ variants are 19.8% and 19.6%, respectively. These relatively high R-factors of the pH 10.7 structures are correlated directly with the relatively poor quality of the diffraction data (See Table 4.7 in Material and methods). The crystals of the complexes were initially grown at pH 6.5 and then transferred to a solution of pH 10.7 prior to data collection. The crystal quality appeared to suffer from the resulting very large change in pH (4.2 units), as the reflections displayed a high mosaicity. An increased stepsize (0.3° between consecutive frames) was used later to collect the data for the Gln¹⁸ variant complex. This proved to be partially effective in overcoming the high mosaicity problem, as the final R-factor of the Gln¹⁸ variant (17.5%) was closer to those of the pH 6.5 structures.

The stereochemical qualities of the structures have been assessed by the program "PROCHECK" (Laskowski et al., 1993). In all the structures, from 84.6% to 89.4% of the residues fall into the most favorably allowed regions of a Ramachandran plot. The only residue that lies in a disallowed region on this plot is Asn100 of SGPB. The conformation of Asn100 is consistent in all the crystal structures of SGPB. The electron density of this residue is well defined in all the structures. One possible reason for the unusual dihedral angles is that Asn100 is located on a sharp turn that is characterized by a cis-proline, Pro99A.

The high quality of each structure can also be appreciated by the representative electron density maps shown in Figures 4.3, 4.7, 4.10 and 4.15.

The overall structure of the complexes

In general there are virtually no conformational changes in either the enzyme or the inhibitor among the various complexes determined at the two extremes of pH. Figure 4.2 is a superposition of all seven of the complexes, showing the main-chain atoms of SGPB and OMTKY3, and all atoms of the P_1 residues. The root mean square (rms) differences of main-chain atoms among these structures are listed in Table 4.2. The observed rms differences of the main-chain atoms are no larger than 0.22 Å (between the Gln¹⁸(N) and Asp¹⁸(H) structures). The rms differences among those structures determined at pH 6.5 are the smallest, ranging from 0.11 Å to 0.15 Å. The experimental error in the coordinates of a well defined, 1.8 Å structure is usually around 0.1 Å. Therefore, the substitution of the P_1 residue does not induce significant structural perturbations to the complexes at the level of the backbone atoms. The overall structures of the complexes are virtually identical among the various complexes.

Table 4.3 presents rms differences in position of common solvent molecules among these complexes. The largest rms difference is 0.47 Å, as observed between the Gln¹⁸(N) and Asp¹⁸(H) structures. Therefore, a positional difference of 1.0 Å (~2 times the rms difference) for a equivalent water molecule in two complexes is considered a reliable indication of migration. The pH 6.5 structures show relatively small rms differences (between 0.31 Å and 0.41 Å) and share a relatively large number of common water molecules. When the pH 10.7 structures were compared with the pH 6.5 structures and among themselves, there are many fewer common water molecules and the rms differences are slightly

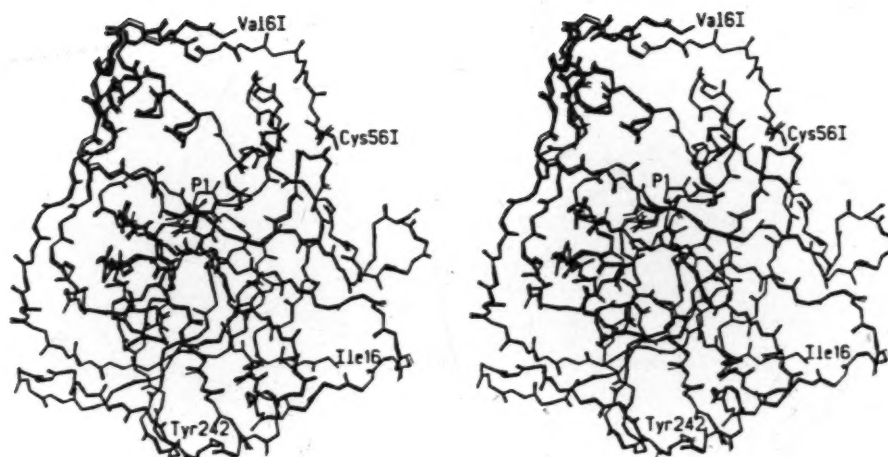


Figure 4.2. Superposition of all the seven complexes. The seven structures are: Glu¹⁸(N), Glu¹⁸(H), Asp¹⁸(N), Asp¹⁸(H), Gln¹⁸(N), Gln¹⁸(H) and Asn¹⁸. The main-chain atoms of SGPB with each OMTKY3 variant and the P₁ residues are shown. The N- and C- termini of SGPB and OMTKY3 are labeled. The remaining six complexes were superposed by least squares onto the Glu¹⁸(N) variant complex.

Table 4.2. RMS differences (Å) of main-chain atoms* among the structures of the various complexes

	Glu ¹⁸ (N)	Glu ¹⁸ (H)	Gln ¹⁸ (N)	Gln ¹⁸ (H)	Asp ¹⁸ (N)	Asp ¹⁸ (H)	Asn ¹⁸
Glu ¹⁸ (N)	-	0.17	0.13	0.18	0.13	0.23	0.12
Glu ¹⁸ (H)		-	0.18	0.19	0.16	0.19	0.16
Gln ¹⁸ (N)			-	0.16	0.15	0.22	0.14
Gln ¹⁸ (H)				-	0.20	0.20	0.19
Asp ¹⁸ (N)					-	0.21	0.11
Asp ¹⁸ (H)						-	0.20
Asn ¹⁸							-

*Number of main-chain atoms used in these comparisons is 944 (740 of SGPB and 204 of OMTKY3).

Table 4.3. RMS differences (Å) of the bound water molecules

	Glu ¹⁸ (N)	Glu ¹⁸ (H)	Gln ¹⁸ (N)	Gln ¹⁸ (H)	Asp ¹⁸ (N)	Asp ¹⁸ (H)	Asn ¹⁸
Glu ¹⁸ (N)	-	0.41[108]	0.34[112]	0.38[92]	0.31[112]	0.46[87]	0.35[117]
Glu ¹⁸ (H)	-	-	0.40[88]	0.42[79]	0.38[86]	0.44[77]	0.42[90]
Gln ¹⁸ (N)	-	-	-	0.40[112]	0.36[109]	0.47[78]	0.35[97]
Gln ¹⁸ (H)	-	-	-	-	0.39[86]	0.46[72]	0.36[78]
Asp ¹⁸ (N)	-	-	-	-	-	0.45[88]	0.34[115]
Asp ¹⁸ (H)	-	-	-	-	-	-	0.46[85]
Asn ¹⁸	-	-	-	-	-	-	-

Numbers in brackets are the number of common water molecules shared by the two structures.

higher. The differences in solvent structures of the complexes may be related to the very large pH change in the crystallization environment. Nevertheless, the majority of the solvent molecules are conserved among these structures at both pHs.

The solvent structure in the S_1 pocket region of each complex, however, differs significantly, each being characteristic of a particular complex at that particular pH. The solvent molecules bound in the S_1 pocket have been compared in a pair-wise manner, Glu¹⁸(N) vs. Glu¹⁸(H), Glu¹⁸(N) vs. Gln¹⁸(N), Glu¹⁸(H) vs. Gln¹⁸(N), Asp¹⁸(N) vs. Asp¹⁸(H), Asp¹⁸(N) vs. Asn¹⁸, Asp¹⁸(H) vs. Asn¹⁸, Glu¹⁸(N) vs. Asp¹⁸(N), Glu¹⁸(H) vs. Asp¹⁸(H), and Gln¹⁸(N) vs. Asn¹⁸. As shown in Table 4.4, the composition and disposition of the bound solvent molecules vary from one complex to another. In general, the solvent molecules in the S_1 pocket are very sensitive to the subtle differences in the polarity of the P_1 residues. There is a rearrangement of solvent structure upon the substitution of the P_1 residue, as will be discussed below.

Structure of the Glu¹⁸ variant at pH 6.5

Figure 4.3 is a representative electron density map, showing the interface between SGPB and the Glu¹⁸ variant of OMTKY3 at pH6.5. There are only three water molecules (W42, W67 and W204) bound in the S_1 pocket; each is secured by forming hydrogen bonds with residues from SGPB and from the inhibitor. The hydrogen-bonding interactions are shown in Figure 4.4. The bonding distances are listed in Table 4.5.

It is unclear if the Glu¹⁸ carboxyl group is ionized in this structure, solely by looking at the associated electron density. The

Table 4.4. Positional differences (Å) of solvent molecules bound in the S₁ pocket (Å)

	W42	W67	W201	W202	W203 (K ⁺)	W204
Glu ¹⁸ (N) vs. Glu ¹⁸ (H)	0.53	0.59	-/+	-/+	-/+	0.64
Asp ¹⁸ (N) vs. Asp ¹⁸ (H)	0.44	0.98*	0.66	0.79	1.26*	1.00*
Glu ¹⁸ (N) vs. Gln ¹⁸ (N)	1.12*	0.51	-/-	-/+	-/-	0.40
Asp ¹⁸ (N) vs. Asn ¹⁸	0.42	1.38*	0.13	1.15*	2.00*	0.13
Glu ¹⁸ (H) vs. Gln ¹⁸ (N)	0.88	0.62	+/-	0.58	+/-	0.68
Asp ¹⁸ (H) vs. Asn ¹⁸	0.44	0.98*	0.66	0.79	1.26*	1.00*
Glu ¹⁸ (N) vs. Asp ¹⁸ (N)	0.13	0.53	-/+	-/+	-/+	1.24*
Glu ¹⁸ (H) vs. Asp ¹⁸ (H)	0.19	0.26	0.62	1.24*	0.32	0.35
Gln ¹⁸ (N) vs. Asn ¹⁸	0.67	0.62	-/+	0.41	-/+	0.81

The presence or absence of a water molecule in the structure is indicated by "+" or "-" respectively. Positional differences that are ~ 1.0 Å or greater (~2 times the maximum observed rms difference) are designated by "*". W203 has been assigned as a K⁺ ion in the Glu¹⁸ (H) and Asp¹⁸ (H) structures.

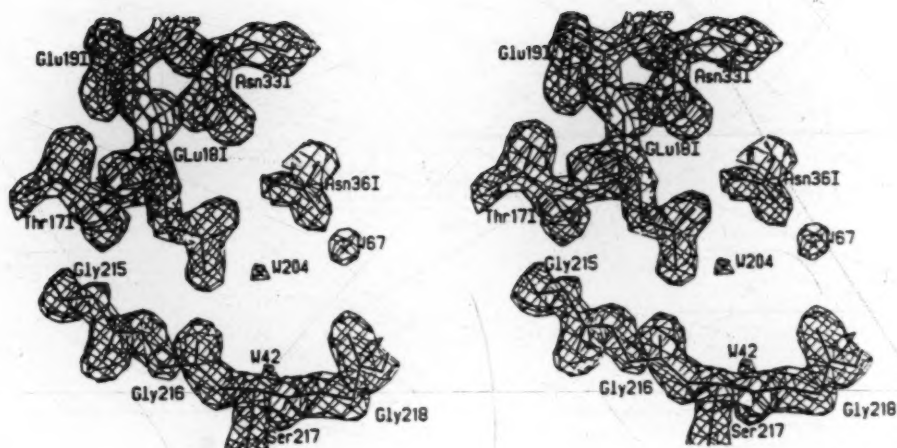


Figure 4.3 A representative portion of the electron density map of the Glu¹⁸(N) structure. The map is centered on part of the interface between SGPB and the Glu¹⁸ variant at pH 6.5. The map was computed with coefficients of $(2m|F_o| - d|F_c|)$ and calculated phases (Read, 1986), and is contoured at the 1.5 σ level.

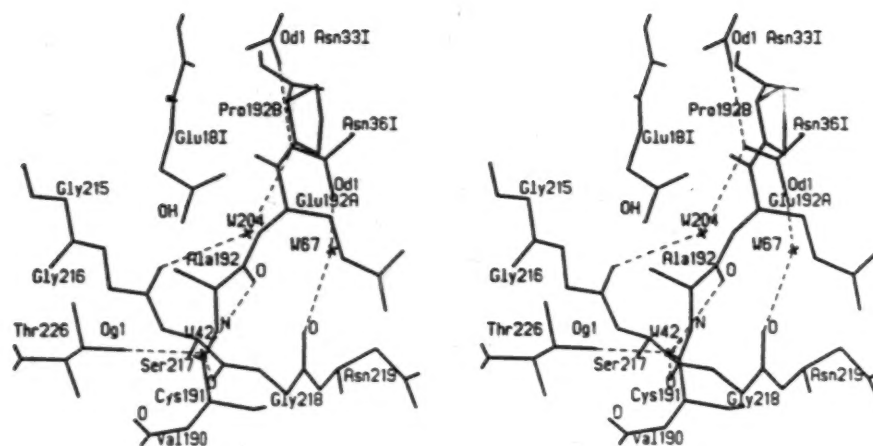


Figure 4.4 The hydrogen-bonding interactions in the Glu¹⁸(N) structure. Dashed lines represent possible hydrogen bonds. The hydrogen-bonding distances are given in Table 4.5.

Table 4.3. Water mediated hydrogen bonds and non-bonded contacts (Å) in the six complexes

H-bond partner	Bonding distance (Å) in complexes					
	Glu18 (W)	Glu18 (E)	Glu18	Asp18 (W)	Asp18 (E)	Asn18
W42						
B-factor (Å ²)	28.3	13.9	15.8	33.9	21.3	27.4
Ala192 O	3.2	3.4	2.9	3.2	3.4	3.1
Ala192 N	2.7	3.2	2.8	2.9	3.4	2.7
Ser217 O	2.9	2.7	3.0	2.9	2.5	3.2
Thr226 OH	3.0	3.4	4.1	3.1	3.2	3.5
W202	-	2.8	3.4	3.4	3.6	3.3
W67						
B-factor (Å ²)	21.2	20.0	35.6	30.4	26.2	37.0
Gly218 O	3.1	3.2	2.7	3.5	3.1	2.7
Asn361 OH	3.3	2.9	3.3	2.9	2.8	3.3
W203	-	2.9 (K ⁺)	-	2.6	3.1 (K ⁺)	3.3
W201						
B-factor (Å ²)	Not present	17.9	Not present	19.3	22.1	22.2
Thr171 O	-	2.8	-	3.8	3.4	3.7
Asn361 N ^{H3}	-	2.8	-	3.0	2.9	3.1
W203 O	-	3.8 (K ⁺)	-	3.0	3.2 (K ⁺)	2.5
W204 O	-	2.8	-	2.9	2.4	2.7
Pl side-chain	-	2.3 (Glu181 O ^{H1})	-	2.6 (Asp181 O ^{H1})	2.4 (Asp181 O ^{H1})	2.7 (Asn181 O ^{H1})
W202						
B-factor (Å ²)	Not present	36.0	18.5	25.7	22.6	20.2
Ala192 O	-	4.0	3.8	3.1	3.7	3.5
Thr226 OH	-	3.1	2.8	4.2	3.5	3.0
W42	-	2.8	3.4	3.4	3.6	3.3
W203	-	3.0 (K ⁺)	-	3.1	2.6 (K ⁺)	2.4
Pl side-chain	-	2.9 (Glu181 O ^{H2})	3.0 (Glu181 N ^{H2})	3.1 (Asp181 O ^{H2})	2.3 (Asp181 O ^{H2})	3.0 (Asn181 N ^{H2})
W203 or K ⁺ ion						
B-factor (Å ²)	Not present	18.3 (K ⁺)	Not present	33.1	27.1 (K ⁺)	49.0
Ala192 O	-	2.6	-	2.5	2.5	3.2
Gly218 O	-	2.9	-	2.8	3.0	4.0
W67	-	2.9	-	2.6	3.1	3.2
W201	-	3.8	-	3.0	3.1	2.5
W202	-	3.0	-	3.1	2.6	2.4
W204	-	2.7	-	3.7	3.0	3.1
Pl side-chain	-	2.7 (Glu181 O ^{H2})	-	4.0 (Asp181 O ^{H1})	3.2 (Asp181 O ^{H2})	3.2 (Asp181 O ^{H1})
W204						
B-factor (Å ²)	32.3	18.6	24.5	32.3	22.6	28.1
Gly216 O	3.3	3.1	3.3	2.8	3.1	2.7
Asn361 OH	3.3	4.2	3.3	4.4	3.4	ND
W201	-	2.8	-	2.9	2.4	2.7
W203	-	2.7 (K ⁺)	-	3.7	3.0 (K ⁺)	3.1

*Residues in brackets are the hydrogen-bonding partners in that specific complex. '-' indicates that the specified hydrogen-bonding partner is absent in that complex.

bond distances for a C=O double bond and a C-O single bond are 1.22 Å and 1.32 Å, respectively, as determined from small molecular structures at very high resolution. The small difference of 0.1 Å is insignificant at 1.8 Å resolution. It is only by inference from the Glu¹⁸(H) structure that the Glu18I carboxyl group is protonated in the Glu¹⁸(N) structure.

The distinction between the -OH and the carbonyl oxygen of the Glu18I side-chain can not be made in the absence of any obviously polarized hydrogen bonds. There are only two atoms, Gly216 N and Pro192B N, that are within the 3.4 Å of hydrogen-bonding distance to the Glu18I carboxyl group. However, Gly216 N donates its proton to Cys16I O as part of an antiparallel β -sheet, and Pro192B N can not form a hydrogen bond. Thus their interactions with the Glu18I carboxyl group are interpreted as normal van der Waals contacts rather than hydrogen bonds. The next nearest neighbor is W204 (3.8 Å), which can not form diagnostic hydrogen bonds with the Glu18I carboxyl oxygen atoms. Therefore, the -OH group and the carbonyl oxygen (C=O) are assigned arbitrarily.

Structure of the Glu¹⁸ variant at pH 10.7

The most striking feature of the Glu¹⁸(H) structure is the presence of a metal ion-associated, electron-dense peak in the S₁ pocket. Figure 4.5 is a difference map computed with coefficients of $(|F_o|_{(H)} - |F_o|_{(N)})$ and phases $\alpha_{cal(N)}$, where $|F_o|_{(H)}$ are the observed structure factor amplitudes of the pH 10.7 structure of the Glu¹⁸ variant complex, $|F_o|_{(N)}$ and $\alpha_{cal(N)}$ are respectively the observed structure factor

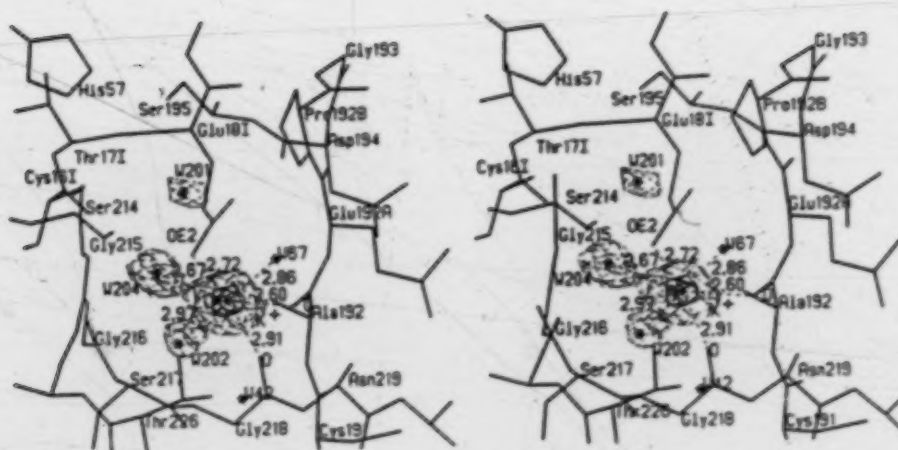


Figure 4.5 The identification of a K^+ ion in the $Glu^{18}(N)$ structure. The map coefficients are $(|F_o|(H) - |F_o|(N))$, and phases, $\alpha_{cal}(N)$, where $|F_o|(H)$ are the observed structure factor amplitudes of the $Glu^{18}(N)$ structure, $|F_o|(N)$ and $\alpha_{cal}(N)$ are respectively the observed structure factor amplitudes and the calculated phases of the $Glu^{18}(N)$ structure. Two contour levels are shown: 5 σ level (shown by dashed lines) and 10 σ level (shown by solid lines).

amplitudes and the calculated phases of the pH 6.5 structure of the Glu¹⁸ variant complex. The highest peak in this difference electron density map is about 10 times higher than the standard deviation of the map, and twice as high as the next highest peak (corresponding to a water molecule). It is a clear indication that the peak corresponds to a metal ion. The metal ion is believed to be a potassium ion, as it displays octahedral coordination and the average K⁺...O distance is 2.79 Å, close to the expected value of 2.8 Å. The identities of the ligands and their distances to the K⁺ ion are given in Table 4.5. The six ligands are Glu18I O²⁻ of the inhibitor, Ala192 O, Gly218 O of SGPB and water molecules W42, W202 and W204.

That a K⁺ ion is present in the Glu¹⁸(H) structure but not in the Glu¹⁸(N) structure makes it immediately clear that the Glu18I carboxyl group is negatively-charged at pH 10.7, and is neutral in the pH 6.5 structure. These results are consistent with the elevated pK_a of Glu18I (8.7) in the complex (Qasim et al, 1995). At pH 10.7, which is 2 pH units above its measured pK_a, the predominant ionic species of the Glu18I carboxyl group should be negatively-charged. Apparently the negative charge buried in the hydrophobic S₁ pocket has attracted a K⁺ ion present in the mother liquor. Potassium and sodium phosphate buffer of 0.4 M were used as precipitant in our crystallization experiments. Except for a shorter ligation distance (2.5 Å as opposed to 2.8 Å of K⁺), a Na⁺ is otherwise identical to a K⁺ ion in its capability of forming octahedral coordination and will compete with a K⁺ for the observed metal ion binding site. The peak height of the top density peak in the difference map correlates with that of a K⁺ ion, and the associated B-factor of this K⁺ is 18.3 Å² in the currently refined

structures. It appears that the observed complex has preferentially bound a K^+ ion, with which the ideal ligation distance can be achieved. But it still is possible that a small population of Na^+ ions are present among K^+ ions in the K^+/Na^+ ion binding site as a result of their competition.

In the experiments for the K_m measurements, sodium ions were present in 500 times excess of the enzyme-inhibitor complex. It is highly likely that a Na^+ ion, in place of the crystallographically observed K^+ ion, was recruited to the S_1 pocket to balance the negatively-charged P_1 side-chains. We believe that the pH 10.7 structure we have observed here closely represents the actual structure present at the high pH end of the observed pH dependent profile of K_m .

Figure 4.6 is a superposition of the Glu¹⁸(N) and the Glu¹⁸(H) structures. The position of the Glu18I carboxyl group is virtually unchanged upon the acquisition of charge. Unlike the Glu¹⁸(N) structure, in which there are no contacts between the Glu18I carboxyl group and water molecules, the ionized Glu18I carboxyl group is heavily hydrated. Accompanying the acquisition of a negative charge on the Glu18I carboxylate group and a counter ion, a K^+ ion, an additional two water molecules, W201 and W202 bind in the S_1 pocket. The three water molecules (W42, W67 and W204) that are present in both the Glu¹⁸(N) and Glu¹⁸(H) structures remain roughly in the same binding sites (Table 4.4). These water molecules form extensive hydrogen-bonds in the interface (Figure 4.6). Residues involved in interacting with these water molecules are Glu18I and Asn36I of the inhibitor, Ala192, Gly216,

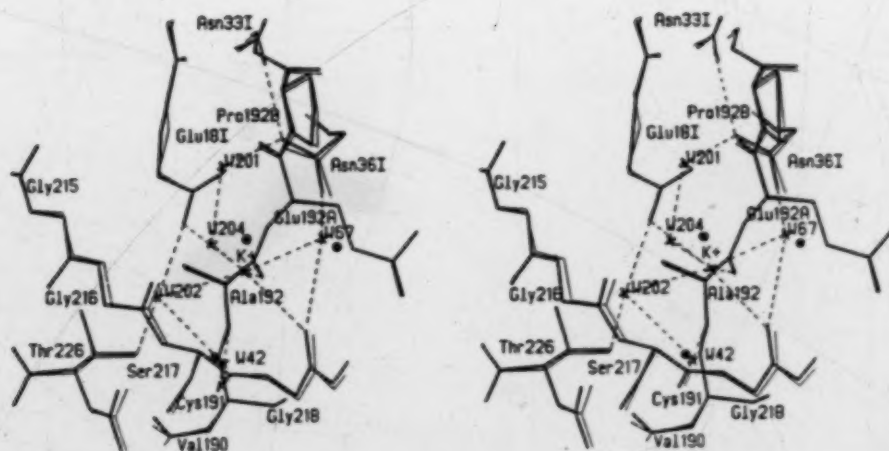


Figure 4.6 Superposition of the Glu¹⁸(H) (thick lines) and the Glu¹⁸(N) (thin lines) structures. Crosses and solid circles represent water molecules observed in the Glu¹⁸(H) and the Glu¹⁸(N) structures, respectively. Dashed lines represent possible hydrogen bonds observed in the Glu¹⁸(H) structure. The hydrogen-bonding distances are shown in Table 4.5.

Ser217, Gly218 and Thr226 of SGPB. The hydrogen bonding distances are listed in Table 4.5.

Another noticeable difference is the coordinated shift of the polypeptide segment flanking Gly216 and Ser217. This segment opens out slightly in the presence of the K⁺ ion and the new water molecules. A small positional change also occurs at the Asn36I sidechain. Asn36I has been shown previously to interact with SGPB via water-bridged hydrogen bonds (Huang et al., 1995). It forms hydrogen bonds with Asn33I, W67, W201 and W204 in this structure (Table 4.6). The small rotation of the Asn36I sidechain is probably related to forming a hydrogen bond with W201.

Structures of the Gln¹⁸ variant at pH 6.5 and pH 10.7

Crystal structures of the complex between SGPB and the Gln¹⁸ variant of OMTKY3 have been determined at pH 6.5 and pH 10.7 as well. The pH 10.7 structure is used as a control experiment, to ensure that the observed K⁺ ion in the pH 10.7 structures of the Glu¹⁸ variant and later of the Asp¹⁸ variant is due solely to the negatively-charged state of the P₁ side-chain. As expected, the structure of the Gln¹⁸ variant at pH 10.7 is virtually identical to that at pH 6.5. The rms difference of main-chain atoms is only 0.16 Å (Table 4.2). Therefore the pH 6.5 structure will be used in the following comparisons to represent the pH insensitive Gln¹⁸ variant structure, since it has been refined to a better crystallographic R-factor due to the small mosaic spread of the reflections from the pH 6.5 crystals.

Table 4.6 Hydrogen bonding and non-bonded interactions (Å) involving Asn331 and Asn361 in the six complexes

H-bond partner	Bonding distance (Å) in complexes					
	Glu18 (M)	Glu18 (N)	Gln18	Asp18 (W)	Asp18 (N)	Asn18
Asn331 N ^{δ2}						
Thr171 O	3.0	3.1	3.1	3.0	3.2	3.0
Glu191 O	2.8	2.8	2.9	3.0	2.8	2.8
Asn331 O ^{δ1}						
Asn361 N	2.8	2.9	2.8	2.8	3.1	2.9
Asn361 N ^{δ2}	3.0	3.0	3.2	3.2	3.3	3.2
Asn361 O ^{δ1}						
W67	3.3	2.9	3.4	2.9	2.8	3.3
W201	Not present	3.9	Not present	2.6	3.7	3.9
W204	3.3	4.2	3.3	3.9	3.4	4.3
Asn361 N ^{δ2}						
W201	Not present	2.8	Not present	3.0	2.9	3.1
W204	3.3	4.3	3.5	4.4	3.7	4.1
P ₁ side-chain	4.3 [Glu181 O ^{δ1}]	4.6 [Glu181 O ^{δ1}]	3.8 [Gln181 O ^{δ1}]	4.9 [Asp181 O ^{δ1}]	4.8 [Asp181 O ^{δ1}]	5.0 [Asn181 O ^{δ2}]

Residues in brackets are the hydrogen-bonding partners in that specific complex.

Figure 4.7 shows the interface between SGPB and the Gln¹⁸ variant of OMTKY3 at pH 6.5. There are four well-ordered water molecules (W42, W67, W202 and W204) bound in the S₁ pocket, forming hydrogen bonds among themselves and with residues of the enzyme and the inhibitor (Table 4.5). The Gln18I side-chain forms hydrogen bonds only with W202 (3.0 Å) and W204 (3.4 Å).

Several important differences have been identified when comparing the Gln¹⁸(N) and the Glu¹⁸(N) structures. Figure 4.8 is a superposition of the pH 6.5 structures of the Gln¹⁸ and Glu¹⁸ variants, overlaid on a difference map. The map coefficients are $(|F_o|_{\text{Gln}} - |F_o|_{\text{Glu}})$ with phases $\alpha_{\text{cal}}(\text{Glu})$, where $|F_o|_{\text{Gln}}$ are the observed structure factor amplitudes of the Gln¹⁸ variant complex, $|F_o|_{\text{Glu}}$ and $\alpha_{\text{cal}}(\text{Glu})$ are respectively the observed structure factor amplitudes and the calculated phases of the pH 6.5 structure of the Glu¹⁸ variant complex. The side-chain of Gln18I, as indicated by the nearby negative peak, occupies a position slightly different from that of the Glu¹⁸(N) structure. In addition, the top two positive peaks in the difference electron density map clearly indicate that W42 adopts a new position, and that a new water molecule, W202, is present in the Gln¹⁸(N) variant structure. The positions of W67 and W204 are very similar to those observed in the Glu¹⁸(N) and the Gln¹⁸(N) structures (Table 4.4).

The current resolution of these structure (1.8Å) can not warrant a clear differentiation between the carboxyl oxygen and the -NH₂ of the Gln18I side-chain. As shown in Figure 4.8, Gln18I N²² is assigned tentatively to a position from where it can form a stronger hydrogen bond with W202 than would Glu18I O⁴¹, as judged by the possible bonding

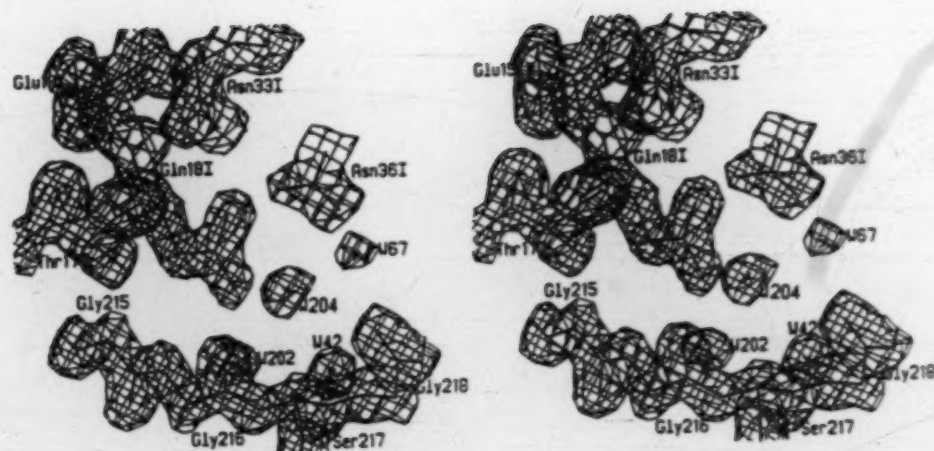


Figure 4.7 A representative portion of electron density map of the Gln¹⁸(N) structure. The map is centered on part of the interface between SGPB and the Gln¹⁸ variant at pH 6.5. The map was computed with coefficients of $(2m|F_o| - d|F_c|)$ and calculated phases, and is contoured at the 1.5σ level.

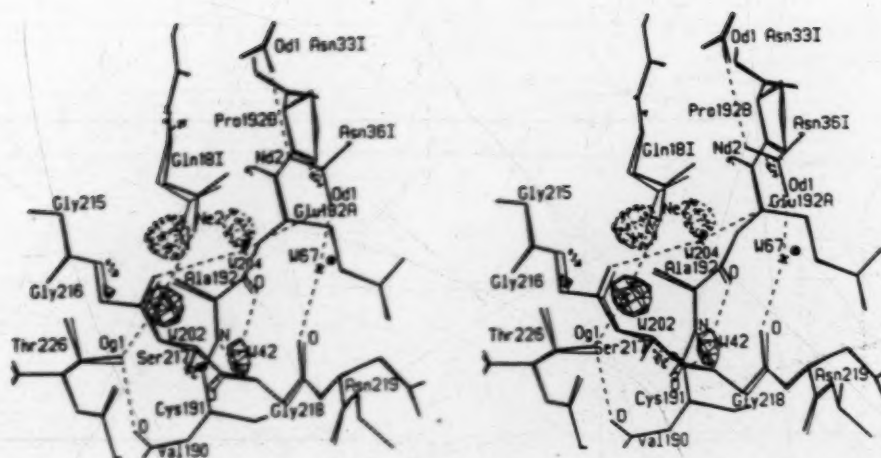


Figure 4.8 Superposition of the Gln¹⁸(N) (thick lines) and the Glu¹⁸(N) (thin lines) structures, overlaid on a difference map. The map coefficients are $(|F_o|(Gln) - |F_o|(Glu))$, phases, $\alpha_{cal}(Glu)$, where $|F_o|(Gln)$ is the observed structure factor amplitudes of the Gln¹⁸(N) structure, $|F_o|(Glu)$ and $\alpha_{cal}(Glu)$ are respectively the observed structure factor amplitudes and the calculated phases of the Glu¹⁸(N) structure. Positive peaks (solid lines, contoured at 5 σ level) and negative peaks (broken lines, contoured at 5 σ level) are shown. Crosses and solid circles represent water molecules observed in the Gln¹⁸(N) and Glu¹⁸(N) structures, respectively. The hydrogen-bonding interactions are listed in Table 4.5.

angle. Nitrogen and oxygen atoms differ by only one electron. It was found that the electron density peak was consistently higher at where the C=O bond was assigned, even when a C-N bond was assigned there (Figure 4.9), indicating that the density at this position is associated with an oxygen atom rather than a nitrogen.

It is noticeable that, in addition to the difference in position of the amide group of Gln18I relative to that of the carboxyl group of Glu18I (Figure 4.8), Gln18I N⁴² and Gly216 N are further apart, when compared with the equivalent distance between Glu18I O⁴² and Gly216 N. This outward shift of the Gly216 to Ser217 segment of polypeptide chain of SGPB may result from the introduction of W202 and the -NH₂ group of Gln18I, which is slightly bulkier than the -OH group of Glu18I.

Structure of the Asp¹⁸ variant at pH 6.5

As shown in Figure 4.10, there are six well-ordered water molecules bound in S₁ pocket at the interface between SGPB and the Asp¹⁸ variant at pH 6.5. In contrast to the pH 6.5 structure of the Glu¹⁸ variant, the Asp18I carboxyl group is well solvated. Both oxygen atoms of the carboxyl group are hydrogen-bonded to water molecules. The six water molecules (W42, W67, W201, W202, W203 and W204) form an extensive hydrogen-bonding network among themselves and with residues of SGPB and the inhibitor (Figure 4.11). The distances of these hydrogen bonds are given in Table 4.5.

As in the case of the Glu¹⁸(N) structure, the Asp18I carboxyl group is protonated in this pH 6.5 structure, by inference of the Asp¹⁸(H)

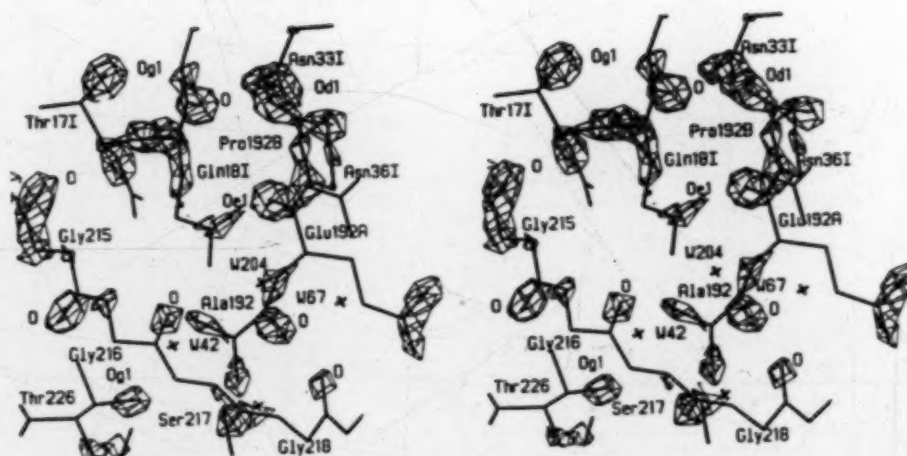


Figure 4.9 Electron density corresponding to the $O^{\delta 1}$ of Gln18 in the Gln¹⁸(N) structure. The map ($2m|F_o| - d|F_c|$ and calculated phases) is contoured at 3.4 σ level.

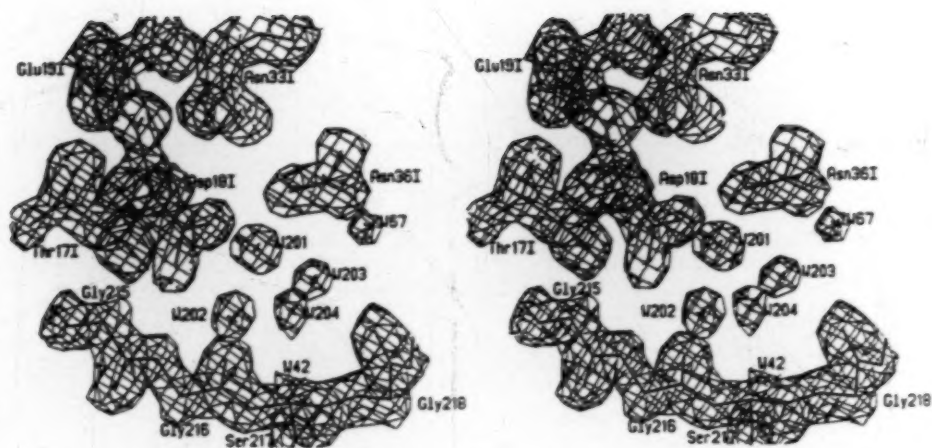


Figure 4.10 A representative portion of electron density map of the Asp¹⁸(N) structure. The map is centered on part of the interface between SGPB and the Asp¹⁸ variant at pH 6.5. The map was computed with coefficients of $(2m|F_o| - d|F_c|)$ and calculated phases, and is contoured at the 1.5σ level.

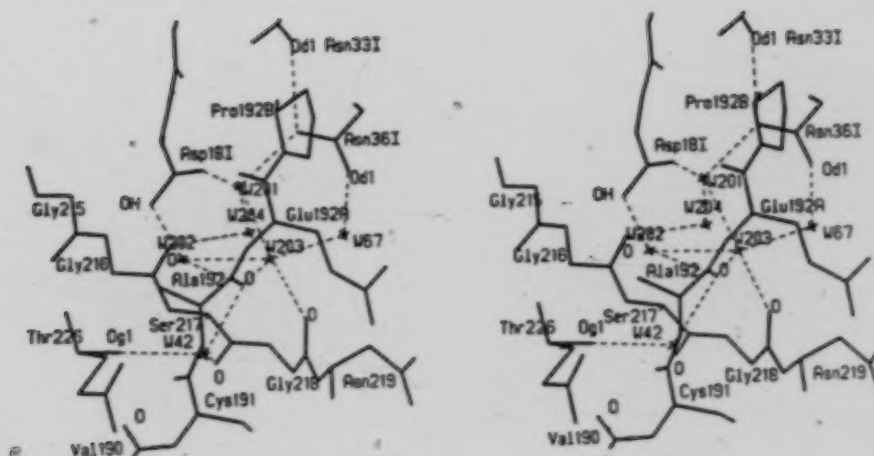


Figure 4.11 Hydrogen-bonding interactions in the Asp¹⁸(N) structure. Dashed lines represent possible hydrogen bonds. The hydrogen-bonding distances are listed in Table 4.5.

structure and the experimentally determined pK_a . There are no obviously polarized hydrogen bonds from which the distinction between the carbonyl carbon and the -OH group can be made. Thus the two carboxyl oxygen atoms are assigned arbitrarily.

Structure of the Asp¹⁸ variant at pH 10.7

A K^+ ion coordinated by the P_1 carboxylate has also identified in the Asp¹⁸(H) structure. Figure 4.12 is a difference map computed with coefficients of $(|F_o|_{(H)} - |F_o|_{(N)})$ and calculated phases $\alpha_{cal}(N)$, where $|F_o|_{(H)}$ are the observed structure factor amplitudes of the pH 10.7 structure of the Asp¹⁸ variant complex, $|F_o|_{(N)}$ and $\alpha_{cal}(N)$ are respectively the observed structure factor amplitudes and the calculated phases of Asp¹⁸(N). Similar to the result for the Glu¹⁸(H) structure, the highest peak (9 σ) in the difference map most probably corresponds to a K^+ ion. As shown in Table 4.5, the K^+ ion is coordinated to seven ligands, three of which are form $K^+ \cdots O$ distances longer than 3 Å. In addition, the coordination geometry adopts an even more distorted octahedron than that observed in the Glu¹⁸(H) structure.

The K^+ ion binding site observed in the Asp¹⁸(H) structure is virtually identical to that in the Glu¹⁸(H) structure (Figure 4.13 and Table 4.4). When the Glu¹⁸(H) and the Asp¹⁸(H) structures were superposed, the positions of the K ion in the two complexes differ by only 0.32 Å. W42, W67 and W204 in the two complexes also superpose very well. W201 and W202, the two water molecules which hydrogen bond to the carboxyl oxygens of the P_1 side-chains, shift slightly in position between the Glu¹⁸(H) and the Asp¹⁸(H) structures. The conformations of

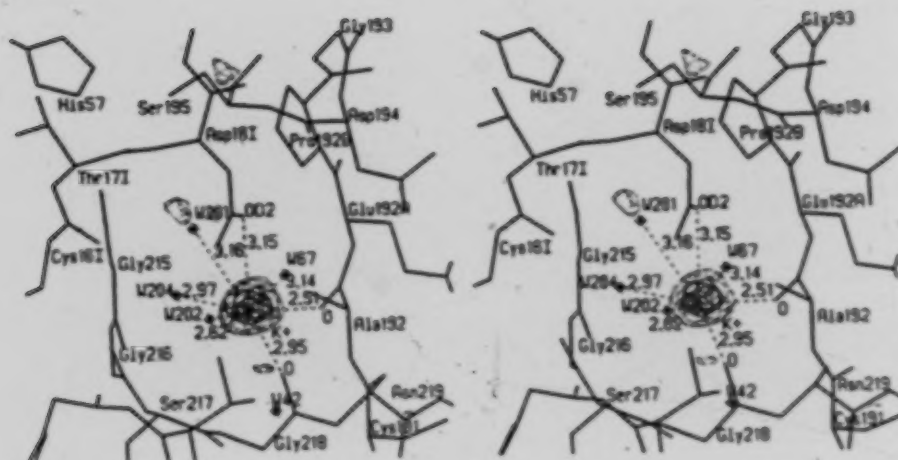


Figure 4.12 The identification of a K^+ ion in the $Asp^{18}(N)$ structure. The map coefficients are $(|F_o|(N) - |F_c|(N))$, phases, $\alpha_{cal}(N)$, where $|F_o|(N)$ are the observed structure factor amplitudes of the $Asp^{18}(N)$ structure, $|F_c|(N)$ and $\alpha_{cal}(N)$ are respectively the observed structure factor amplitudes and the calculated phases of the $Asp^{18}(N)$ structure. Two contour levels are shown: 4σ level (shown by dashed lines) and 8σ level (shown by solid lines).

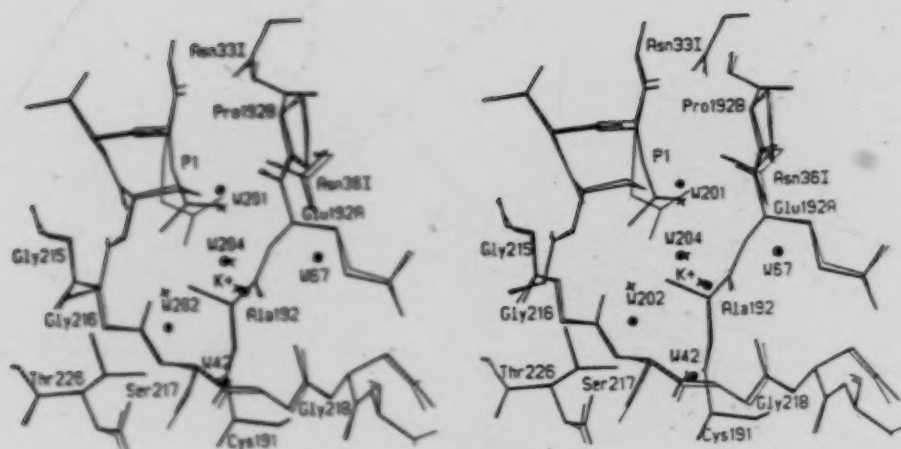


Figure 4.13 Superposition of the Asp¹⁸(H) (shown in thick lines) and the Glu¹⁸(H) (shown in thin lines) structures. Crosses and solid circles represent solvent molecules observed in the Asp¹⁸(H) and Glu¹⁸(H) structures, respectively. The possible hydrogen bonds are shown by dashed lines

the polypeptide segment flanking Gly215 and Gly216 are remarkably similar in the two structures, so is that of Asn36I.

The presence of a cation in the S_1 pocket is a contributing diagnostic of the negatively-charged state of the Asp18I carboxylate group in this Asp¹⁸(H) structure. By inference, the Asp18I carboxyl group is believed to be protonated in the pH 6.5 structure. These observations support the pK_a measurement data for the Asp¹⁸ variant of OMTKY3, in which the pK_a of the Asp18I carboxyl group was found to be 9.4 in association with SGPB (Qasim & Laskowski, unpublished data). The major ionic species of the Asp18I carboxyl group at pH 10.7 is negatively-charged, and will inevitably attract a positive charge to the S_1 pocket in order to balance the charge. In the solution in which the pK_a data were obtained, the most likely ion to be recruited as a counter ion are sodium ions. The resulting structure, including the metal-ligation geometry, is believed to be similar to our crystallographically observed, K^+ ion bound complex. The formation of a salt bridge between the Na^+ and the Asp18I carboxylate will compensate, to a certain extent, the very large energy cost to bury a single charge in the hydrophobic S_1 pocket.

The total number of solvent molecules bound in the S_1 pocket are identical in the Asp¹⁸(H) and Asp¹⁸(N) structures (Figure 4.14 and Table 4.4). But there are considerable positional adjustments of these water molecules, particularly the concerted shifts of W67 (0.98 Å), W203 (1.26 Å) and W204 (1.00 Å) towards the Asp18I carboxylate. The Asp18I and Asn36I side-chains show small rotation relative to those in the Asp¹⁸(N)

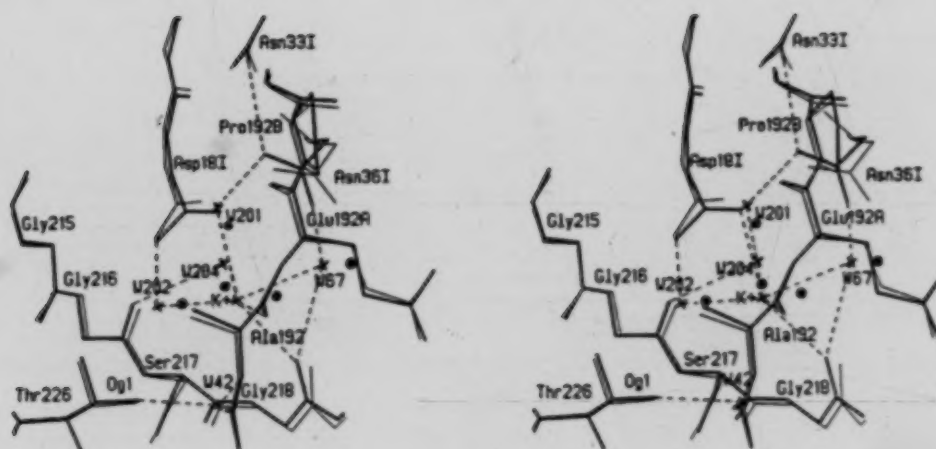


Figure 4.14 Superposition of the Asp¹⁸(H) (thick lines) and the Asp¹⁸(N) (thin lines) structures. Crosses and solid circles represent water molecules observed in the Asp¹⁸(H) and Asp¹⁸(N) structures, respectively. Dashed lines represent possible hydrogen bonds observed in the Asp¹⁸(H) structure. The hydrogen bonding distances are listed in Table 4.5.

structure, probably due to the rearrangement of the solvent structure in the S_1 pocket.

Structure of the Asn¹⁸ variant at pH 6.5

Figure 4.15 shows the electron density map at the molecular interface in the complex between SGPB and the Asn¹⁸ variant at pH 6.5. There are five water molecules bound in the S_1 pocket, with a disposition that is different from any of the other structures (Table 4.4). Their positions relative to those observed in the Asp¹⁸(N) structure are shown in Figure 4.16. Overlaid on Figure 4.15 is a difference map with coefficients $(|F_o|_{(Asn)} - |F_o|_{(Asp)})$ and phases $\alpha_{cal}(Asp)$, where $|F_o|_{(Asn)}$ are the observed structure factor amplitudes of the Asn¹⁸ variant complex, $|F_o|_{(Asp)}$ and $\alpha_{cal}(Asp)$ are, respectively, the observed structure factor amplitudes and the calculated phases of the Asp¹⁸(N) variant complex. The most noticeable features of this map are the peaks which suggest that W202 shifts from its position in the Asp¹⁸(N) structure by 1.2 Å to a new position in the Asn¹⁸ variant structure. Coupled to this shift of W202, W203 and W67 shift towards the same direction, by a distance of 1.4 Å and 2.0 Å respectively (Table 4.4). Consequently, the hydrogen-bonding partners and bonding distances involving the three water molecules are changed (Table 4.5). The rearrangement of these solvent molecules is obviously related to the substitution of the carboxyl group by an amide group at the P_1 side-chains. Glu181 N²² is assigned tentatively to a position from where a hydrogen bond is made to W202. W202 forms hydrogen bonds with Asp181 N³² (3.0 Å) and Thr226 O⁷¹ (3.0 Å) in the Asn¹⁸ structure, and similarly,

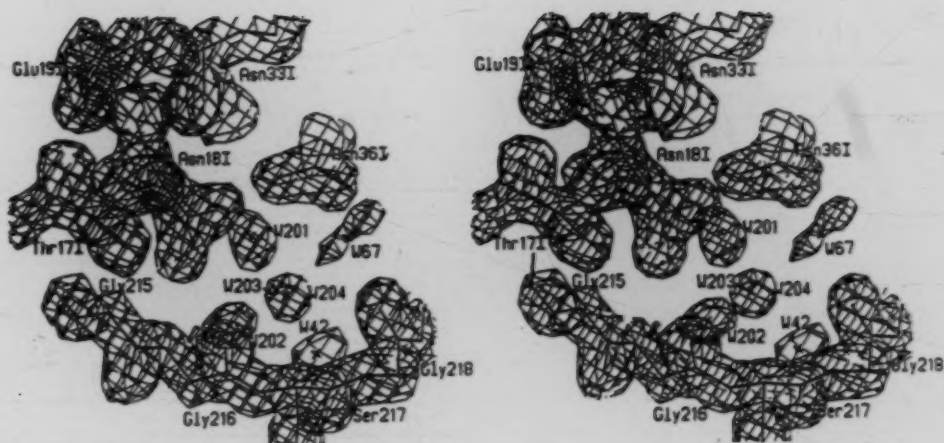


Figure 4.15 A representative portion of the electron density map of the Asn¹⁸ structure. The map is centered on part of the interface between SGPB and the Asn¹⁸ variant. The map was computed with coefficients of $2m|F_o| - d|F_c|$ and calculated phases, and is contoured at the 1.5σ level.

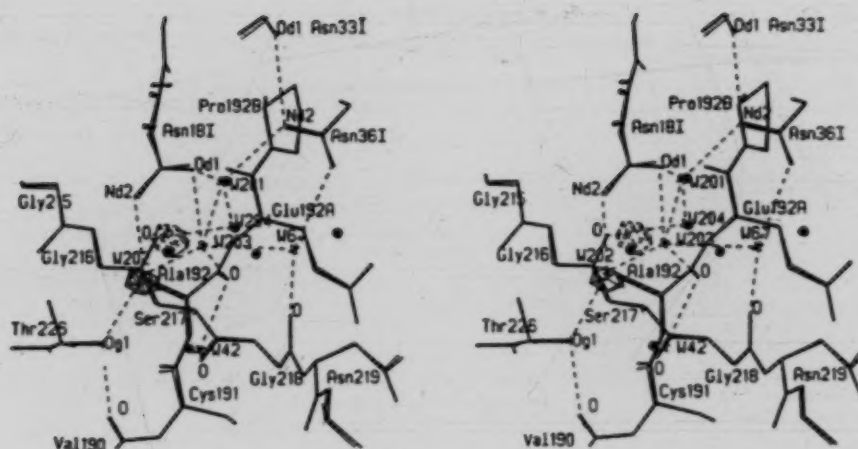


Figure 4.16 Superposition of the Asn¹⁸(N) (thick lines) and the Asp¹⁸(N) (thin lines) structures, overlaid on a difference map. The map coefficients are $(|F_o|(\text{Asn}) - |F_o|(\text{Asp}))$, phases, $\alpha_{\text{cal}}(\text{Asp})$, where $|F_o|(\text{Asn})$ is the observed structure factor amplitudes of the Gln¹⁸(N) structure, $|F_o|(\text{Asp})$ and $\alpha_{\text{cal}}(\text{Asp})$ are respectively the observed structure factor amplitudes and the calculated phases of the Asp¹⁸(N) structure. Positive peaks (solid lines, contoured at 4 σ level) and negative peaks (broken lines, contoured at 4 σ level) are shown. Crosses and solid circles are water molecules observed in the Asn¹⁸(N) and Asp¹⁸(N) structures. The hydrogen-bonding distances are given in Table 4.5.

with Glu18I N²² (2.8 Å) and Thr226 O²¹ (3.0 Å) in the Glu¹⁸(N) structure. Such water bridged hydrogen bonds between Thr226 O²¹ and the P₁ side-chains are not observed in any of the other four structures.

DISCUSSION

When comparing the Glu¹⁸(N) and Glu¹⁸(H), Asp¹⁸(N) and Asp¹⁸(H), the most striking difference is the acquisition of a K⁺ ion in the S₁ pocket in those structures determined at pH 10.7. We have taken the presence of a cation as diagnostic of the negatively-charged state of the P₁ side-chain carboxylate groups in the two structures. By inference, the P₁ side-chain carboxyl groups are believed to be protonated in the respective complexes with SGPB determined at pH 6.5.

The formation of a salt bridge between a bound K⁺/Na⁺ ion and a negatively-charged carboxylate oxygen of P₁ side-chains following the deprotonation of Glu18I and Asp18I has given us a direct measurement of the energetic effects of introducing an ion-pair to the hydrophobic S₁ pocket. Qasim et al. have shown that the K_d is about $1.0 \times 10^{10} \text{ M}^{-1}$ for SGPB interacting with the Glu¹⁸ variant of OMTKY3 when Glu18I is protonated in both the free and complexed states; the K_d is $6.0 \times 10^5 \text{ M}^{-1}$ when Glu18I is negatively-charged (Qasim et al., 1995). The reduction in the stability of the enzyme-inhibitor complex, which results from the deprotonation of the P₁ side-chains and the subsequent formation of the salt bridge, is about 5.6 Kcal/mol, as estimated from the reduction in K_d. In other words, the energy cost to accommodate the K⁺...carboxylate

salt bridge in the highly hydrophobic S_1 pocket of SGPB is about 5.6 Kcal/mol.

It had been shown by a continuum electrostatic analysis that salt bridges buried in the interior of proteins are destabilizing, when compared with hydrophobic interactions (Mendsch & Tidor, 1994). This study has suggested that the destabilizing effect arises from the very large energy cost of desolvating ions, of which even the formation of favorable salt bridges between ions can only compensated partly, but not all. It was calculated that, when salt bridging side-chains are mutated to hydrophobic isosteres, the average free energy change would be 3.5 Kcal/mol. Subsequently this number has been determined experimentally in a study of the Arc repressor dimerization (Waldburger et al., 1995). A salt-bridged triad is buried in the interface of this dimeric protein. The entire salt-bridge network can be replaced by hydrophobic residues and the dimer becomes additionally stabilized by up to 4.5 Kcal/mol. By measuring the $\Delta\Delta G^\circ$ of introducing an ion-pair to a preformed hydrophobic pocket, our study provides an important, direct approach to test the destabilizing effect of salt bridges. One advantage of our approach is that there is no substitution-induced structural perturbation, which is a major complication of many mutagenesis studies.

A major question that we have asked is: "When the Glu18I side-chain is ionized, how does the Glu¹⁸ variant manage to remain associated with SGPB with a moderately strong K_d of $6 \times 10^5 \text{ M}^{-1}$? It is highly energetically unfavorable to transfer a charge from aqueous solution into a hydrophobic environment. One would expect that introducing an

anion to the hydrophobic S_1 pocket would dissociate the complex between SGPB and the inhibitor, in a fashion similar to the pH induced unfolding of proteins. The Glu¹⁸(H) and Asp¹⁸(H) structures have revealed that a counter ion K^+ is recruited to balance the negative charge on the P_1 carboxyl groups. The K^+ ...carboxylate salt bridge thus stabilizes the protein complex to a certain extent. The experimental measurements of the association constants were carried out in the presence of excessive amount of sodium ions. A much weaker K_a would be anticipated when the enzyme-inhibitor complex is deprived of monovalent metal ion. In the absence of metal ions, the difference in stability between the complex of SGPB with negatively-charged Glu18I and the complex of SGPB with protonated Glu18I should provide an estimate of the energy cost to insulate a negative charge in the S_1 pocket. In a mutagenesis study of T4 lysozyme, it has found that the protein thermostability was reduced by 5.5 to 6.9 Kcal/mol, when an Asp or a Lys was introduced to positions in the hydrophobic core (Dao-pin et al., 1991).

It is very interesting that when the Glu18I carboxyl group is protonated in the free and in the complexed states, the Glu¹⁸ variant of OMTKY3 is actually the third best inhibitor towards SGPB, next to the Leu¹⁸ variant ($K_a = 5.6 \times 10^{10} M^{-1}$) and the Met¹⁸ variant ($K_a = 2.7 \times 10^{10} M^{-1}$), and is 14 times better than the Gln¹⁸ variant. The protonated Glu18I appears to be more hydrophobic than Gln18I, which is consistent with its longer retention time in reversed-phase HPLC (Guo et al., 1986). The crystal structures of Glu¹⁸(N) and Gln¹⁸(N) reveal that the interactions involving Glu18I and Gln18I are different. In the Glu¹⁸(N) structure, the carboxyl group of Glu18I forms van der Waals contacts directly with Gly216 O and Pro192B O of SGPB. There are 4 ordered water

molecules in the Gln¹⁸(N) structure, as opposed to 3 in the Glu¹⁸(N) structure. Whereas Gln18I O⁶¹ is in van der Waals contacts with Pro192B O, Gln18I N⁶² makes no direct contact with the enzyme, but forms a hydrogen bond with W202 (Fig. 4.8). The entropic cost for ordering the additional water (W202) in the S₁ pocket is estimated to be about 1.7 Kcal/mol, ($\Delta S = 6RT/2$, assuming that W202 loses all the six degrees of translational and rotational entropy), which is approximately equal to the 1.6 Kcal/mol reduction in stability of the complex. However, it may be an oversimplification to attribute the reduction in stability of the complex of Gln¹⁸(N) solely to the entropic cost of ordering W202. The reason for this is that first, the hydrophobicities of Gln18I and protonated Glu18I are different, so are the resulted contacts involving them. Secondly, the positions of W42 differ by 1.1 Å when the Gln¹⁸(N) and Glu¹⁸(N) complexes are compared (Table 4.4), consequently the hydrogen bonding partners, angles and distances are changed (Table 4.5).

The carboxyl group of Asp18I in the Asp¹⁸ variant of OMTKY3 displays a higher pK_a than that of Glu18I in the Glu¹⁸ variant. This might be explained in part by the highly distorted coordination geometry of the K⁺ ion in the Asp¹⁸(H) structure. The average K⁺...O ligation distance in the Asp¹⁸(H) structure is 2.93 Å, as opposed to that of 2.78 Å in the Glu¹⁸(H) structure. The salt bridge distance between the K⁺ ion and Asp18I O⁶² is 3.2 Å, as compared to that of 2.7 Å in the Glu¹⁸(H) structure (Figure 4.16). This distance (3.2Å) is ~0.4 Å longer than the normal K⁺...O distance of 2.8 Å. Consequently the driving force towards deprotonation, conferred by the K⁺ ion, is significantly reduced. In the experimental K_a measurement, where the cation is a Na⁺ ion, whose ionic radius is even smaller than that of a K⁺ ion by 0.3 Å, the

elongated salt bridge will have considerably less influence on the deprotonation equilibrium.

When measured at pH 8.3, the Gln¹⁸ to Glu¹⁸ substitution results in a loss of 3.3 Kcal/mol in binding free energy, the Asn¹⁸ to Asp¹⁸ a loss of 2.2 Kcal/mol. Because Asp18I has a higher pK_a than that of Glu18I, one expects the negatively-charged population is only 7.4% in the Asp¹⁸ case, in comparison to 28.6% for Glu18I. This may be part of the reason that Gln¹⁸ to Glu¹⁸ substitution is more destructive than the Asn¹⁸ to Asp¹⁸ substitution at pH 8.3. When the P₁ residue is fully protonated (at very low pH), the Glu¹⁸ variant should bind SGPB more tightly than the Asp¹⁸ variant.

The side-chains of protonated glutamic acid, ionized glutamic acid and glutamine are isosteric. In all three cases, burying the P₁ side-chains in the S₁ pocket of SGPB does not introduce significant structural perturbations to the main-chain atoms of either the enzyme or the inhibitor. The precise conformations of these P₁ side-chains differ only slightly. However, the total number, the composition and the disposition of the solvent molecules bound in the S₁ pocket are remarkably different from one structure to another (Table 4.4). Whereas there are only three water molecules in the Glu¹⁸(N) structure, there are four water molecules and a K⁺ ion in the Glu¹⁸(H), and four water molecules in the Gln¹⁸(N) structure. The hydrogen-bonding interactions involving the solvent molecules are not all similar among the three structures (Table 4.5). The same observation holds true for the Asp¹⁸(N), Asp¹⁸(H) and Asn¹⁸ structures (Table 4.4 and 4.5). It appears that the solvent molecules bound in the S₁ pocket are very sensitive to

the stereochemistry and hydrophobicity of the P₁ side-chain. It has been shown in our previous study of hydrophobic P₁ side-chains that the solvent molecules bound in the S₁ pocket undergo rearrangement, in response to the size of the P₁ side-chain (Huang et al., 1995). The comparisons among the Glu¹⁸(N), Glu¹⁸(H), Gln¹⁸(N), Asp¹⁸(N), Asp¹⁸(H) and Asn¹⁸(N) structures suggest that the rearrangement also takes place in response to the charged state and polarity of the P₁ side-chain. The characteristic solvent organization in the S₁ pocket of each complex provides a magnified image of the polarity of the S₁ pocket.

Asn36I has been shown to be involved in water-bridged hydrogen bonds with residues encompassing the S₁ pocket (Huang et al., 1995). The interactions between Asn36I and bound water molecules (Table 4.6) have been further illustrated by the observed adjustment of Asn36I to solvent rearrangement during the deprotonation process of Glu18I of Asp18I. When comparing the Glu¹⁸(N) and Glu¹⁸(H) structures, or the Asp¹⁸(N) and Asp¹⁸(H) structures, it is observed that small positional adjustments of Asn36I occur, probably coupled to the changes in solvent structure in the S₁ pocket.

Theoretical calculations of pK_a from atomic models have in many cases yielded good agreement with experimental results (reviewed by Honig & Nicholls, 1995). However, it will be a challenging task to calculate the pK_a of Glu18I and Asp18I when complexed with SGPB. Firstly, without knowing precedents, it will be difficult to predict the acquisition of a K⁺ ion (or a Na⁺ ion) accompanying the ionization of the P₁ side-chain. Secondly, the K⁺ ion is not an integral part of the protonated structure. It is drawn into the complex at a certain stage

of the deprotonation process and in turn presumably influences the equilibrium of deprotonation.

MATERIAL AND METHODS

Crystallization

SGPB, purified according to the procedure of Jurasek et al. (1979), was kindly provided by Dr. L. B. Smillie. The cloning, overexpression system and the purification procedures for variants of OMTKY3 have been described (Lu et al, 1992; Lu, 1994). The 1:1 stoichiometric complexes of SGPB and OMTKY3 variants were prepared by mixing the two in a 1:1.05 molar ratio of enzyme to inhibitor. Crystals of the complex were grown by hanging drop methods, under conditions similar to those previously reported (Fujinaga et al, 1982). Each drop contained 3 μ l of protein solution (10mg/ml) and 3 μ l of sodium/potassium phosphate buffer (0.4 M, pH6.5). For the pH 10.7 structure, crystals were transferred to a pH 10.7 sodium/potassium phosphate solution and soaked for at least 24 hours prior to the data collection. These crystals are isomorphous to those of the complex of SGPB with native OMTKY3. Unit cell parameters of the crystals of the complexes are listed in Table 4.7.

Data collection

X-ray diffraction data for the seven OMTKY3 variants complexed with SGPB were collected on twin San Diego Multiwire System area detectors. Graphite monochromated Cu K α radiation was generated by a Rigaku rotating anode generator RU-200BH operating at 40 KV and 150mA. An 0.15 degree ω stepsize between consecutive frames was used to

Table 4.7. Statistics of diffraction data

Space Group	Glu ¹⁸ (N)		Glu ¹⁸ (H)		Gln ¹⁸ (N)		Gln ¹⁸ (H)		Asp ¹⁸ (N)		Asp ¹⁸ (H)		Asn ¹⁸	
	P2 ₁	P2 ₁	P2 ₁	P2 ₁	P2 ₁	P2 ₁	P2 ₁	P2 ₁	P2 ₁	P2 ₁	P2 ₁	P2 ₁	P2 ₁	P2 ₁
Unit Cell Parameters														
a (Å)	45.54	45.67	45.41	45.35	45.55	45.59	45.58							
b (Å)	54.68	54.84	54.70	54.65	54.66	54.61	54.67							
c (Å)	45.58	45.74	45.59	45.62	45.63	45.62	45.61							
β (°)	119.19	119.32	119.17	119.20	119.13	119.24	119.17							
Maximum resolution (Å)	1.8	1.8	1.8	1.8	1.8	1.8	1.8							
No. of crystals used	1	1	1	1	1	1	1							
Total No. of observations	40,784	36,566	36,831	35,022	35,657	31,119	46,014							
No. of unique reflections	17,019	17,051	17,167	17,153	17,068	17,193	17,283							
Average redundancy	2.4	2.1	2.1	2.0	2.1	1.8	2.7							
Completeness of data overall (%)	91.2	89.2	90.4	90.0	88.0	89.7	89.2							
1.9 Å - 1.8 Å (%)	85.1	88.3	89.5	89.2	80.1	86.4	85.0							
R _{merge} (%)	5.33	11.75	8.94	8.75	6.38	10.49	7.01							

$$R_{\text{merge}} = \frac{\sum_{hkl} (|I_1| - \langle I \rangle)}{\sum_{hkl} I_1}$$

collect the data from the pH 6.5 crystals. A 0.2 degree ω stepsize was used for the pH 10.7 crystals of the Glu¹⁸ and Asp¹⁸ variants. Later a 0.3° ω stepsize was used for the pH 10.7 crystal of the Gln¹⁸ variants. A complete data set was usually collected from a single crystal. The data-collection time for each complex lasted about 2 days, and the ambient temperature was 22° during collection. The data-collection statistics for all seven complexes are given in Table 4.7.

Structure determination and refinement

The procedures for the determination and refinement of the seven structures are the same as previously reported (Huang et al., 1995). The initial phases for each of the complexes of SGPB with the OMTKY3 variants were taken from those of the complex of SGPB with native OMTKY3 (Read et al., 1983). A $2|F_o| - |F_c| \exp(i\alpha_{cal})$ map, where $|F_o|$ are the observed structure-factor amplitudes of the variant complexes, $|F_c|$ and α_{cal} are the structure-factor amplitudes and phases calculated from the structure of Read, et al. (1983), was calculated with the program SIGMAA (Read, 1986) and displayed with a locally modified version of FRODO (Jones, 1986). The P₁ residue, Leu18I, was replaced by the appropriate P₁ side-chains to fit best the observed electron density in these maps. After energy minimization and simulated annealing with X-PLOR (Brunger, 1986), each of the models was subjected to TNT refinement (Tronrud, 1992). All reflections with $1.8 \text{ \AA} < d_{min} < 20.0 \text{ \AA}$ were used for the refinements. After approximately ten cycles of TNT refinement, a new $2|F_o| - |F_c|$ map was computed for visual inspection of the fit of the model to the density, and for selecting solvent peaks to add to the molecular model of the asymmetric unit. Special attention was given to

regions where there were noticeable conformational disparities with the previously reported structure and where individual atoms displayed high thermal factors. Up to this stage no water molecules had been built into the complex model. Additionally, an $(|F_o| - |F_c|) \exp(i\alpha_{cal})$ difference electron density map was made to assist in assigning water molecules. Water molecules that had been present in the previously reported structure (Read et al., 1983) were built into the OMTKY3 variant complex model if there were associated densities for them above the 2.5σ contour level in the difference map. The manually-adjusted model with the added solvent peaks as water molecules was then subjected to another 10 cycles of TNT refinement. Afterwards water molecules with B factors higher than $50 \text{ \AA}^2$ were examined closely. A water molecule would be deleted if there was no associated density at the 1σ contour level on a $2|F_o| - |F_c|$ map and if the B factor was greater than $60 \text{ \AA}^2$. New water molecules were added if there was electron density above the 2.5σ level on a $|F_o| - |F_c|$ difference map. The geometry of potential hydrogen bond donor or acceptor atoms was also checked to assure that it was meaningful to introduce a water molecule in that region. Refinement was judged to reach convergence typically after a total of 30 cycles of TNT refinement, when the mean positional shift was less than $0.015 \text{ \AA}$.

Structure comparisons

Superpositions of the complex structures were done in a pairwise manner using the *lsq* commands in the program "O" (Jones et al, 1991). The root-mean-square (rms) differences for common atoms were obtained subsequently from the program.

REFERENCES

- Ardelt WA & Laskowski M Jr. 1991. Effect of single amino acid replacements on the thermodynamics of the reactive site peptide bond hydrolysis in ovomucoid third domain. *J Mol Biol* 220:1041-1053.
- Bigler TL, Lu W, Park SJ, Tashiro M, Wieczorek WR & Laskowski M Jr. 1993. Binding of amino acid side chains to preformed cavities: Interaction of serine proteinases with turkey ovomucoid third domains with coded and noncoded P₁ residues. *Protein Sci* 2:786-799.
- Bode W, Wei AZ, Huber R, Meyer E, Travis J & Neumann S. 1986. X-ray crystal structure of the complex of human leukocyte elastase (PMN elastase) and the third domain of the turkey ovomucoid inhibitor. *EMBO J* 5:2453-2458.
- Bode W & Huber R. 1992. Natural protein proteinase inhibitors and their interactions with proteinases. *Eur J Biochem* 204:433-451.
- Dao-pin S, Anderson DE, Baase WA, Dahlquist FW and Matthews BW. 1991. Structural and thermodynamic consequences of burying a charged residue within the hydrophobic core of T4 lysozyme. *Biochemistry* 30:11521-11529.
- Fujinaga M, Sielecki AR, Read RJ, Ardelt W, Laskowski M Jr. & James MNG. 1987. Crystal and molecular structures of the complex of α -chymotrypsin with its inhibitor turkey ovomucoid third domain at 1.8 Å resolution. *J Mol Biol* 195:397-418.
- Guo D, Mant CT, Taneja AK, Parker JMR & Hodges RS. 1986. Prediction of peptide retention times in reversed-phase high-performance liquid chromatography: I. determination of retention coefficients of amino acid residues of model synthetic peptides. *J Chromatography* 359:499-517.
- Hendsch ZS & Tidor B. 1994. Do salt bridges stabilize proteins? a continuum electrostatic analysis. *Protein Sci* 3:211-226.

- Honig B & Nicholls A. 1995. Classical electrostatics in biology and chemistry. *Science* 268:1144-1149.
- Huang K, Lu W, Anderson S, Laskowski M Jr. & James MNG. 1987. Water molecules participate in proteinase-inhibitor interactions: Crystal structures of Leu¹⁸, Ala¹⁸ and Gly¹⁸ variants of Turkey ovomucoid inhibitor third domain complexed with *Streptomyces griseus* proteinase B. *Protein Sci* (in press)
- James MNG, Delbaere LTJ & Brayer GD 1978. *Can J Biochem* 56:423-438
- Jones TA. 1985. Interactive computer graphics: PRODO. *Meth Enzym* 115:157-171. With modifications for TOM UQV/CIT v.2.8.0 by Camillau C, Israel M, Griffin G. & Chirino A.
- Jones TA, Zou JY, Cowan SW, Kjeldgaard M. 1991. Improved methods for building protein models in electron density maps and the location of errors in these models. *Acta Cryst A* 47:110-119.
- Jurasek L, Johnson P, Olafson RW & Smillie LB. 1971. An improved fractionation system for pronase on CM-sephadex. *Can J Biochem* 49:1195-1201.
- Jurasek L, Olafson RW, Johnson P & Smillie LB. 1977. Relationship between the structures and activities of some microbial serine proteinases. I. Purification, enzymatic properties and primary structures of *Streptomyces griseus* protease A, B and trypsin. In *Miami winter symposium, Vol. II. Proteolysis and physiological regulation*. Ribbons DW & Brew K, eds. New York: Academic press. pp 93-123.
- Laskowski M Jr. & Kato I. 1980. Protein inhibitors of proteinases. *Biochemistry* 49:593-626.
- Laskowski RA, MacArthur MW, Moss DS & Thornton JM. 1993. PROCHECK: a program to check the stereochemical quality of protein structures. *J Appl Cryst* 26:283-291.

- Lu W. 1994. Energetics of the interactions of ovomucoid third domain variants with different serine proteinases. *Ph.D. Thesis*, Purdue University, West Lafayette, Indiana.
- Qasim MA, Ranjbar MR, Wynn R, Anderson S & Laskowski M Jr. 1995. Ionizable P₁ residues in serine proteinase inhibitors undergo large pK_a shifts on complex formation. *J Biol Chem* (in press).
- Read RJ, Fujinaga M, Sielecki AR & James MNG. 1983. Structure of the complex of *Streptomyces griseus* protease B and the third domain of the turkey ovomucoid inhibitor at 1.8 Å resolution. *Biochemistry* 22:4420-4433.
- Read RJ & James MNG. 1986. Introduction to the protein inhibitors: X-ray crystallography. In *Proteinase inhibitors* Barrett, AJ and Salvesen G eds. Amsterdam:Elsevier. pp. 301-336.
- Read RJ. 1986. Improved Fourier coefficients for maps using phases from partial structures with errors. *Acta Cryst* A42:140-149.
- Rupley JA, Butler L, Gerring M, Hartdegen FJ & Pecoraro R. 1967. Studies on the enzymatic activity of lysozyme, III, the binding of Saccharides. *Pro Nat Acad Sci USA* 57:1088-1092.
- Schechter I & Berger A. 1967. On the size of the active site in proteases. I. Papain. *Biochem Biophys Res Commun* 27:157-162.
- Tronrud DE. 1992. Conjugate-direction minimization: An improved method for the refinement of macromolecules. *Acta Cryst* A48:912-916.
- Waldburger CD, Schildbach JF & Sauer RT. 1995. Are buried salt bridges important for protein stability and conformational specificity?. *Nature Struct Biol* 2:122-128.
- Yang A-S, Honig B. 1992. Electrostatic effects on protein stability. *Curr Opin in Struct Biol* 2:40-45.

CHAPTER 5

CONCLUSIONS

This thesis has reported eleven crystal structures of eight different intermolecular complexes, the complex of porcine pancreatic elastase with a 63 amino acid residue inhibitor from *Ascaris*, and the complexes of SGPB with seven P₁ OMTKY3 variants.

The crystal structure of the complex between PPE and a *Ascaris* chymotrypsin/elastase inhibitor has been determined at 2.4 Å resolution, by the molecular replacement method coupled with the electron density averaging techniques. The 63 residue polypeptide chain of the *Ascaris* C/E-1 inhibitor displays a novel fold that has not been observed among any other protein inhibitors of known structure (Figure 2.1a). The disc-shaped molecule is characterized by five disulfide bridges and two small β -sheets. The fact that this inhibitor does not have a well-defined hydrophobic core is understandable, given that nearly 75% of the constitutive residues are hydrophilic. An interesting feature of the structure is the presence of two salt-bridge triads on either face of the molecule (Figure 2.6 and 2.7). One salt-bridge triad is formed by Glu9I, Glu18I and Arg48I, the other is formed by Glu39I, Ser41I and Arg44I. Their roles are possibly to provide additional stability to the protein. It is realized that these six residues are absolutely conserved among the five *Ascaris* chymotrypsin/elastase inhibitors, and four of

them, Glu9I, Glu19I, Glu39I and Arg48I, are also present in the *Ascaris* trypsin inhibitor ATI (Figure 2.5). The hydrogen bond between Arg44I and Ser41I is likely to be replaced by the hydrophobic interactions between Ala44I and Ile41I at the equivalent positions in the *Ascaris* trypsin inhibitor. It is thus tempting to propose mutagenesis studies to test the possible roles of these salt bridges in providing stability and folding specificity.

The determined crystal structure of the C/E-1 inhibitor has negated the disulfide bridge pairing proposed by chemical methods (Bernard et al., 1994). The topology of the disulfide bridges is shared by the ATI, whose NMR structure was determined nearly at the same time as our crystal structure (Grasberger et al., 1994). Hydrophobic interactions are believed to be the driving force in protein folding. Thus it is curious how, in the absence of a hydrophobic core, the C/E-1 inhibitor assumes the correct linkage of disulfide bridges, and eventually folds into a compact molecule.

The interactions between elastase and the inhibitor have been examined and compared with other protein inhibitors in detail (Table 2.2 and Figure 2.9). In addition to the common properties shared by other canonical inhibitors, a distinguishing feature of the C/E-1 inhibitor is an extended secondary binding region. Up to 20 residues from the inhibitor are involved in interacting with elastase. A pore has been identified in the C/E-1 inhibitor, which embraces the Arg217A side-chain (Figure 2.10).

The C/E-1 inhibitor is the first member of its family whose crystal structure has been determined. The structure of the complex has

revealed for the first time how PPE interacts with a canonical protein inhibitor.

Crystal structures have been determined for the complexes of SGPB with a series of recombinant variants of OMTKY3: Leu¹⁸, Ala¹⁸, Gly¹⁸, Glu¹⁸, Gln¹⁸, Asp¹⁸, Asn¹⁸. Thermodynamic studies (measurements of association constants) have been carried out in Dr. M. Laskowski's laboratory to characterize the reactivities of these P_i variants towards six cognate enzymes (Lu, 1994). The crystal structures of these complexes have been analyzed and compared, in an attempt to relate the measured thermodynamic properties to the observed structural details.

Firstly, the structure of the Leu¹⁸ variant, the wild-type recombinant OMTKY3 (residues 5-56), has reassured that it is essentially identical in conformation to the naturally-occurring OMTKY3 (residue 1-56) in the reactive-site loop region (Figure 3.2).

The variations in the P_i side-chains do not introduce major conformational changes to the complex. The typical rms differences of the main-chain atoms of SGPB and OMTKY3 variants are around 0.15 Å for well-refined structures, and are no greater than 0.22 Å in the worst case (Table 4.2). The segment flanking Gly215 and Gly216 has been identified to be the most sensitive region to the variations in the P_i side-chains. This segment of SGPB shifts by up to 0.8 Å towards the side-chains of Leu18I and the protonated Glu18I, resulting in a interface that is more compact than that in the other complexes and in the native SGPB (Figure 3.5 and 4.6). This P_i-induced shift is particularly interesting given that the Leu¹⁸ and protonated Glu¹⁸ are

respectively the best and the third best inhibitors among all the 20 coded P_1 variants.

Another dynamic element is the solvent structure in the S_1 binding pocket region (Table 3.3 and 4.4). Comparisons among these structures and of each with native, uncomplexed SGPB have revealed that each complex features a unique solvent structure in the S_1 pocket. Specifically, the total number, the composition and the relative positions of these solvent molecules vary according to the size, the shape and the chemical nature of the P_1 side-chains. In other words, solvent molecules experience redistributions in adjusting to the binding of different P_1 side-chains. In this sense, the solvent molecules bound in the interface have become an extended structural element of the complexes, mediating extensive interactions between the enzyme and the inhibitors (Table 3.4 and 4.5).

The ionizable P_1 residues Glu18I and Asp18I of OMTKY3 variants display unusually high pK_a values (8.7 and 9.4 respectively) when complexed with *Streptomyces griseus* Proteinase B. The complexes remain associated even after the carboxyl group of the P_1 residues are ionized (Qasim et al., 1995). The 1.8 Å crystal structures of the Glu¹⁸ and Asp¹⁸ variants of OMTKY3 have been determined in complex with SGPB at pH 6.5 (the neutral form) and pH 10.7 (the ionized form). The crystal structures of the Gln¹⁸ and Asn¹⁸ variants of OMTKY3 have been determined at pH 6.5 (the Gln¹⁸ variant also at pH 10.7). Surprisingly, a potassium ion has been identified in the pH 10.7 structures of the Glu¹⁸ and Asp¹⁸ variants (Figure 4.5 and 4.10). Apparently it has been recruited to the S_1 pocket to balance the negative charge of the P_1

residue. This K^+ ion forms a salt bridge with the carboxylate group of the P_1 side-chains.

The crystallographically observed recruitment of a K^+ ion is very likely parallel to the true event that took place during the pH-dependent K_d measurements. There were adequate amounts of Na^+ ions present in the enzyme stock solution (at least 500 times excess) used for K_d measurement. Except for a shorter ligation distance (2.5 Å for Na^+ as opposed to 2.8 Å of K^+), a Na^+ is otherwise identical to a K^+ ion in its capability of forming octahedral coordination. Thus it is reasonable to believe that, during the measurement of the variation of K_d as a function of pH, a Na^+ ion was recruited to form a salt bridge with the negatively-charged P_1 carboxyl group, at pH values above the pK_a of the P_1 side-chains. The energy cost of burying this salt bridge is approximately 5.6 Kcal/mol, as estimated from the pH-dependent increase in the free energies of binding.

It has been commonly accepted that a buried charge group destabilizes proteins (Dao-pin et al., 1991). On the other hand, buried ion-pairs (salt bridges) were believed to stabilize proteins, based primarily on studies on chymotrypsin (Fersht, 1972), and on T4 lysozyme (Anderson et al., 1990). It has been realized only recently that salt bridges buried in the interior of a protein are in fact destabilizing; the protein stability would be increased if residues participating in such buried salt bridges can be replaced with hydrophobic isosteres (Hendsch & Tidor, 1994; Waldburger et al., 1995). Our studies on the Glu¹⁸ variant have led to a direct measurement of the destabilizing effect of a buried salt bridge on the stability of the complex, without

involving mutagenesis. The absolute magnitude of the destabilizing effect estimated from our study is slightly larger than that derived from theoretical calculations (Hendsch & Tidor, 1994) and that obtained from mutagenesis study on Arc repressor protein (Waldburger et al., 1995), possibly due to the additional entropic cost to recruit a Na^+ and two additional water molecules.

In addition to the crystallographically determined complexes, the complexes of SGPB with four non-coded P_1 variants, Abu, Ape, Ahx and Ahp, have been modeled on the basis of the Ala^{18} variant structure. A homologous series of straight aliphatic side-chain (from 0 to 5 carbon atoms) P_1 variants, Gly^{18} , Ala^{18} , Abu^{18} , Ape^{18} , Ahx^{18} and Ahp^{18} , have provided a good model to evaluate the contribution of hydrophobic interactions to the free energy of binding. The simplicity of this selected hydrophobic series is that, as observed when modeling the complexes, no conformational strains would occur, either to the enzyme or the inhibitor, upon the insertion of the P_1 side-chain to the S_1 pocket. For this set of homologous variants, a linear correlation has been found between the change in free energy of binding and the hydrophobic surface area buried in the enzyme-inhibitor interface. The correlation constant gives an estimate of the hydrophobic effect of $34 \text{ cal mol}^{-1} \text{ \AA}^{-2}$. This value is somewhat lower than that estimated from cavity-creating studies of barnase, suggesting that the interface between SGPB and OMTKY3 is not as hydrophobic as the interior of barnase (Kellis et al., 1989). This is likely to be true given the observation that the enzyme-inhibitor interface is partially hydrated with ordered water molecules. The hydrophobic effect estimated from our study should be one of the most reliable estimations, favored by the fact that the

binding of OMTKY3 to SGPB invokes no substantial conformational changes that commonly occur in and complicate the interpretation of the cavity-creating studies in other protein systems.

For this hydrophobic series of variants, the thermodynamic property, K_d , can be linked to the structural property, the amount of hydrophobic surface area that is buried upon the complex formation. In other words, the value of K_d of a complex can be described by the degree of complementarity between the P_1 side-chain and the S_1 pocket. The association of serine proteinase-protein inhibitors is believed to be entropy driven, where the dominant force is the favorable entropic contribution, which is due to the tendency of two hydrophobic, complementary surfaces coming to interact and displacing orderly bound water molecules from each other's surface. Measurements of the enthalpic and entropic contributions to the free energy of binding show a low value for ΔH (0 to +5 Kcal/mol), whereas the value of ΔS is always positive and high (+30 to +60 entropy units), for several systems investigated (eglin^c with human leukocyte elastase, α -chymotrypsin and subtilisin Carlsberg, Ascenzi et al., 1988a; BPTI with human α , β and γ -thrombin, Ascenzi et al., 1988b). By calorimetric measurements of the complexes between SGPB and OMTKY3, one would be able to dissect the enthalpic and entropic contributions to the free energy of binding. It is important to know if ΔS of the binding process is a linear function of the hydrophobic surface area buried upon complex formation, for the selected series of straight aliphatic side-chain P_1 variants. The knowledge of the entropic and enthalpic contributions to the free energy of binding will shed light on understanding why the Glu¹⁸ variant with a protonated P_1 carboxyl group, a hydrophilic group, displays an

association constant and a binding mode (van der Waals contacts) resembling that of hydrophobic groups (Figure 4.4).

REFERENCES

- Anderson DE, Becktel WJ & Dahlquist FW. 1990. pH-induced denaturation of proteins: a single salt bridge contributes 3-5 kcal/mol to the free energy of folding of T4 lysozyme. *Biochemistry* 29:2403-2408.
- Ascenzi P, Amiconi G, Menegatti E, Guarneri M, Bolognesi M & Schnebli HP. 1988a. Binding of the recombinant proteinase inhibitor eglin c from leech *Hirudo medicinalis* to human leukocyte elastase, bovine α -chymotrypsin and subtilisin Carlsberg: thermodynamic study. *J Enz Inhibit* 2:167-172.
- Ascenzi P, Coletta M, Amiconi G, De Cristofaro R, Bolognesi M, Guarneri M & Menegatti E. 1988b. Binding of the bovine basic pancreatic trypsin inhibitor (Kunitz) to human α , β and γ -thrombin: a kinetic and thermodynamic study. *Biochim Biophys Acta* 956:156-161.
- Bernard, V. D. & Peanasky, R. J. (1993). The serine protease inhibitor family from *Ascaris suum*: Chemical determination of the five disulfide bridges. *Arch. Biochem. Biophys.* 303, 367--376.
- Dao-pin S, Anderson DE, Baase WA, Dahlquist FW & Matthews BW. 1991. *Biochemistry* 30:11521-11529.
- Ferth AR. 1972. Conformational equilibria in α - and δ -chymotrypsin. The energetics and importance of the salt bridge. *J Mol Biol* 64:497-509.
- Grasberger BL, Clore GM & Gronenborn AM. 1994. High-resolution structure of *Ascaris* trypsin inhibitor in solution: direct evidence for a pH-

induced conformational transition in the reactive site. *Structure* 2:669-678.

Hendsch ZS & Tidor B. 1994. Do salt bridges stabilize proteins? a continuum electrostatic analysis. *Protein Sci* 3:211-226.

Lu W. 1994. Energetics of the interactions of ovomucoid third domain variants with different serine proteinases. *Ph.D. Thesis*, Purdue University, West Lafayette, Indiana.

Kellis JT Jr, Nyberg K & Fersht AR. 1989. Energetics of complementary side-chain packing in a protein hydrophobic core. *Biochemistry* 28:4914-4922.

Qasim MA, Ranjbar MR, Wynn R, Anderson S & Laskowski M Jr. 1995. Ionizable P₁ residues in serine proteinase inhibitors undergo large pK_a shifts on complex formation. *J Biol Chem* (in press).

Waldburger CD, Schildbach JF & Sauer RT. 1995. Are buried salt bridges important for protein stability and conformational specificity?. *Nature Struct Biol* 2:122-128.